Doctoral Thesis

Harmonic Analysis of Deep Convolutional Neural Networks

Author(s):
Wiatowski, Thomas

Publication Date:
2018

Permanent Link:

https://doi.org/10.3929/ethz-b-000252349 →

HARMONIC ANALYSIS OF DEEP CONVOLUTIONAL NEURAL NETWORKS

DISS. ETH NO. 24556

HARMONIC ANALYSIS OF DEEP CONVOLUTIONAL NEURAL NETWORKS

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH
(Dr. sc. ETH Zurich)

presented by

THOMAS WIATOWSKI

M.Sc. in Mathematics, TU Munich

born on 20.12.1987
citizen of Germany

accepted on the recommendation of

Prof. Dr. Helmut Bölcskei, examiner
Prof. Dr. Thomas Hofmann, coexaminer

2018

für meine Eltern, Anna und Norbert

Abstract

A central task in machine learning, computer vision, and signal processing is to extract characteristic features of signals. Feature extractors based on deep convolutional neural networks (DCNNs) have been applied with significant success in a wide range of practical machine learning tasks such as classification of images in the ImageNet data set (Krizhevsky et al., 2012), image captioning (Vinyals et al., 2015), or control-policy-learning to play Atari games (Mnih et al., 2015) or the board game Go (Silver et al., 2016). Since DCNN architectures lead to remarkable results across a broad range of applications, it is essential to understand their underlying mechanisms. In this thesis, we develop a mathematical theory of DCNNs for feature extraction using concepts from applied harmonic analysis. We investigate the impact of DCNN topology and building blocks—convolution filters, non-linearities, and pooling operators—on the network's feature extraction capabilities.

The mathematical analysis of feature extractors generated by DCNNs was initiated by Mallat in (Mallat, 2012). Specifically, (Mallat, 2012) analyzed so-called scattering networks, where signals are propagated through layers that employ directional wavelet filters and modulus non-linearities but no intra-layer pooling. The resulting wavelet-modulus feature extractor is horizontally (i.e., in every network layer) translation-invariant (where the wavelet scale parameter determines the amount of invariance) and stable with respect to (w.r.t.) certain non-linear deformations, both properties of significance in practical feature extraction applications.

In the first part of this thesis, we complement Mallat's results by

developing a theory of DCNNs for feature extraction encompassing general convolutional transforms, or in more technical parlance, general semi-discrete frames (including Weyl-Heisenberg, curvelet, shearlet, ridgelet, and wavelet frames), general Lipschitz-continuous non-linearities (e.g., rectified linear units, shifted logistic sigmoids, hyperbolic tangents, and modulus functions), and general Lipschitz-continuous pooling operators emulating sub-sampling and averaging. In addition, all of these elements can be different in different network layers. For the resulting network (called *generalized* scattering network) we prove a translation invariance result which is of vertical nature in the sense of the network depth determining the amount of invariance, and we establish deformation sensitivity bounds that apply to signal classes with inherent deformation insensitivity such as, e.g., band-limited functions, cartoon functions (Donoho, 2001) (which provide a good model for natural images), and Lipschitz functions. The essence of our results is that vertical (i.e., asymptotically in the network depth) translation invariance and limited sensitivity to non-linear deformations are guaranteed by the network structure per se rather than the specific convolution filters, non-linearities, and pooling operators.

In the second part of this thesis, we study the DCNN topology, specifically the depth and width. Many practical machine learning tasks employ *very* deep convolutional neural networks (He et al., 2015). Such large depths pose formidable computational challenges in training and operating the network. It is therefore important to understand how fast the energy contained in the propagated signals (a.k.a. feature maps) decays across layers. In addition, it is desirable that the feature extractor generated by the network be informative in the sense of the only signal mapping to the all-zeros feature vector being the zero input signal. This "trivial null-set" property can be accomplished by asking for "energy conservation" in the sense of the energy in the feature vector being proportional to that of the corresponding input signal. We address these questions for the class of scattering networks that employ the modulus non-linearity, no pooling, and general filters that are allowed to be different in different

network layers. We establish conditions for energy conservation (and thus for a trivial null-set) and characterize corresponding feature map energy decay rates. Specifically, we find that under mild analyticity and high-pass conditions on the filters (which encompass, inter alia, various constructions of Weyl-Heisenberg filters, wavelets, ridgelets, (α)-curvelets, and shearlets) the feature map energy decays at least polynomially fast. For broad families of wavelets and Weyl-Heisenberg filters, the guaranteed decay rate is shown to be exponential. Moreover, we provide handy estimates of the number of layers needed to have at least $((1-\varepsilon)\cdot 100)\%$ of the input signal energy be contained in the feature vector.

In the third and final part of this thesis, we focus on the practically relevant discrete-time case, introduce new DCNN architectures, and propose a mathematical framework for their analysis. We establish deformation and translation sensitivity results of local and global nature, and we investigate how certain structural properties of the input signal are reflected in the corresponding feature vectors. Our theory applies to general filters and general Lipschitz-continuous non-linearities and pooling operators. Experiments on handwritten digit classification and facial landmark detection—including a feature importance evaluation—complement the theoretical findings.

Kurzfassung

Das Extrahieren von charakteristischen Merkmalen aus Signalen ist ein wichtiges Problem im maschinellen Lernen und Sehen sowie in der Signalverarbeitung. Feature extractors[1], die auf tiefen neuronalen Faltungsnetzwerken (TNFNs) basieren, werden mit grossem Erfolg in vielen Bereichen des maschinellen Lernens angewandt. Beispiele dafür sind die Klassifizierung von Bildern des ImageNet Datensatzes (Krizhevsky et al., 2012), das Generieren von Bildbeschreibungen (Vinyals et al., 2015) oder das Lernen von Strategien, die es ermöglichen, Computerspiele (Mnih et al., 2015) oder das Brettspiel „Go“ (Silver et al., 2016) zu spielen. Aufgrund des breiten Anwendungsspektrums und der bemerkenswerten Erfolge ist es von zentraler Bedeutung, diejenigen Mechanismen zu verstehen, die der Netzwerkarchitektur zugrunde liegen. In dieser Dissertation entwickeln wir eine mathematische Theorie für das Extrahieren von features mittels TNFNs, die auf Konzepten der angewandten harmonischen Analyse basiert. Wir untersuchen dabei, wie die TNFN-Topologie und -Bausteine—Filter, Nichtlinearitäten und pooling[2] Operatoren—die Fähigkeit der Netzwerke, charakteristische features aus Signalen zu extrahieren, beeinflussen.

Die mathematische Analyse von feature extractors, die auf TNFNs basieren, wurde in (Mallat, 2012) initiiert. Mallat analysierte sogenannte scattering networks[3], in denen Signale durch Netzwerkschich-

[1] Auf Deutsch: Merkmalsextraktoren.
[2] Auf Deutsch: Bündelung.
[3] Auf Deutsch: Streunetzwerke.

ten, die wavelet Filter und modulus Nichtlinearitäten verwenden, jedoch auf pooling Operatoren verzichten, propagiert werden. Der korrespondierende wavelet-modulus feature extractor ist horizontal (d.h. in jeder Netzwerkschicht) translationsinvariant (wobei der Grad der Invarianz durch den wavelet Skalierungsparameter bestimmt wird) und stabil gegenüber gewissen nichtlinearen Deformationen.

Im ersten Teil dieser Dissertation ergänzen wir die Ergebnisse von Mallat, indem wir eine Theorie für das Extrahieren von features mittels TNFNs entwickeln, die es ermöglicht, i) allgemeine semi-diskrete frames (z.B. Weyl-Heisenberg, curvelet, shearlet, ridgelet und wavelet frames), ii) allgemeine Lipschitz-stetige Nichtlinearitäten (z.B. rectified linear units, logistic sigmoids, hyperbolic tangents und den modulus) und iii) allgemeine Lipschitz-stetige pooling Operatoren (z.B. sub-sampling oder averaging) zu verwenden. Ferner ist es möglich, unterschiedliche Filter, Nichtlinearitäten und pooling Operatoren in unterschiedlichen Netzwerkschichten zu benutzen. Für diese Architekturen, die wir *generalized* scattering networks nennen, beweisen wir ein Translationsinvarianz-Resultat, das von vertikaler Natur ist (d.h. die Netzwerktiefe bestimmt den Grad der Invarianz), und wir leiten Deformationssensibilitäts-Garantien her, die für Signalklassen gelten, die inhärent insensibel gegenüber Deformationen sind. Beispiele für solche Signalklassen sind bandbegrenzte Funktionen, Lipschitz-stetige Funktionen sowie cartoon functions (Donoho, 2001), welche sich gut dazu eignen, Bilder zu modellieren. Die Essenz unserer Ergebnisse ist, dass vertikale (d.h. in der Netzwerktiefe asymptotische) Translationsinvarianz und Insensibilität gegenüber nichtlinearen Deformationen durch die Netzwerkstruktur an sich gewährleistet sind, und nicht durch die spezifische Wahl der Filter, Nichtlinearitäten und pooling Operatoren.

Im zweiten Teil dieser Dissertation untersuchen wir die TNFN-Topologie, insbesondere die Tiefe und Breite der Netzwerke. In vielen Anwendungsbereichen des maschinellen Lernens werden *sehr* tiefe neuronale Faltungsnetzwerke vewendet (He et al., 2015). Solche grossen Netzwerktiefen bereiten sowohl im Training als auch in der Anwendung der Netzwerke rechentechnische Probleme. Es ist daher von

zentraler Bedeutung zu verstehen, wie schnell die Energie, die in den feature maps[4] enthalten ist, mit zunehmender Netzwerktiefe abfällt. Ferner ist es wünschenswert, dass das einzige Signal, das durch den feature extractor auf den Null-Vektor abgebildet wird, das Null-Eingangssignal ist. Diese „triviale Nullmengen“ Eigenschaft gilt, wenn die Energie des feature vectors proportional zu der Energie des Eingangssignals ist. Konkret untersuchen wir das Abfallen der feature map Energie und die Erhaltung der feature vector Energie für scattering networks, welche allgemeine Filter, die modulus Nichtlinearität und keine pooling Operatoren verwenden. Wir leiten Bedingungen für Energieerhaltung (und damit für eine triviale Nullmenge) her und charakterisieren die Energieabklingraten der feature maps. Wir zeigen, dass unter Analytizitäts- und Hochpassbedingungen an die Filter (die z.B. von gewissen Weyl-Heisenberg Filtern, wavelets, ridgelets, (α)-curvelets und shearlets erfüllt werden) die Energie mindestens polynomiell in der Netzwerktiefe abfällt. Für einige Familien von wavelet und Weyl-Heisenberg Filtern beweisen wir, dass die Abklingrate sogar exponentiell in der Netzwerktiefe ist. Unsere Energieabkling-Resultate ermöglichen es uns, diejenige Netzwerktiefe zu spezifizieren, die benötigt wird, damit mindestens $((1-\varepsilon)\cdot 100)\%$ der Eingangssignalenergie im feature vector enthalten ist.

Im dritten und letzten Teil dieser Dissertation betrachten wir den praktisch relevanten zeitdiskreten Fall, präsentieren neue TNFN-Architekturen und stellen die mathematischen Grundlagen vor, die für deren Analyse notwendig sind. Wir beweisen Resultate zur Deformations- und Translationssensibilität des feature extractors, die von lokaler und globaler Natur sind, und wir untersuchen, wie sich bestimmte strukturelle Eigenschaften des Eingangssignals im feature vector widerspiegeln. Die von uns entwickelte Theorie kann auf Netzwerke angewandt werden, die allgemeine Filter, allgemeine Lipschitz-stetige Nichtlinearitäten und allgemeine Lipschitz-stetige pooling Operatoren verwenden. Experimente zur Klassifizierung von handgeschriebenen

[4] Auf Deutsch: Propagierte Signale.

Ziffern und zur Erkennung von Gesichtspartien ergänzen die theoretischen Ergebnisse.

Acknowledgements

I would like to express my deepest gratitude to my supervisor, Prof. Helmut Bölcskei, for his excellent guidance, support, vision, and encouragements throughout the last four years. Helmut, you gave me the academic freedom I wanted, and the guidance I needed. Thank you.

Thank you Prof. Philipp Grohs, Aleksandar Stanić, Michael Tschannen, and Verner Vlačić for many inspiring discussions and contributions which were valuable for the outcome of this thesis.

I would like to thank Prof. Helmut Bölcskei and Prof. Thomas Hofmann for acting as examiners for this thesis.

Thanks go to Dr. Céline Aubel, Recep Gül, Dr. Erwin Riegler, Dr. David Stotz, and Michael Tschannen for the enjoyable times we shared.

Thanks go also to ETH Zurich for providing a world-class learning, research, and working environment.

Zu guter Letzt danke ich meinen Eltern, Anna und Norbert, meiner Schwester Patricia, und Laetitia für eure grenzenlose Liebe, Unterstützung und Geduld. Dziękuję, Danke, Merci.

Contents

CHAPTER 1

Introduction

Deep convolutional neural networks (DCNNs) have led to breakthrough results in numerous practical machine learning tasks (Rumelhart et al., 1986; LeCun et al., 1990, 1998, 2010, 2015; Krizhevsky et al., 2012; Bengio et al., 2013; He et al., 2015; Mnih et al., 2015; Goodfellow et al., 2016; Silver et al., 2016). While DCNNs can be used to perform classification (or other machine learning tasks such as, e.g., control-policy-learning to play Atari games (Mnih et al., 2015) or the board game Go (Silver et al., 2016)) directly, typically based on the output of the last network layer, they can also act as stand-alone feature extractors (Serre et al., 2005; Huang and LeCun, 2006; Mutch and Lowe, 2006; Ranzato et al., 2006, 2007; Pinto et al., 2008; Jarrett et al., 2009) with the resulting features fed into a classifier such as a support vector machine (SVM) (Cortes and Vapnik, 1995). The present thesis pertains to the latter philosophy and develops a mathematical theory of DCNNs for *feature extraction*.

1.1. DEEP CONVOLUTIONAL FEATURE EXTRACTION: ARCHITECTURES, INVARIANCES, AND DEFORMATION SENSITIVITY (CHAPTER 3)

A central task in machine learning is feature extraction (Duda et al., 2001; Bishop, 2009; Bengio et al., 2013) as, e.g., in the context of handwritten digit classification (LeCun and Cortes, 1998). The features to be extracted in this case correspond, for example, to the edges of the digits. The idea behind feature extraction is that feeding characteristic features of the signals—rather than the signals themselves—to a classifier (such as, e.g., a SVM) improves classification performance. Specifically, non-linear feature extractors can map input signal space dichotomies that are not linearly separable into linearly separable feature space dichotomies (Bishop, 2009). Sticking to the example of handwritten digit classification, we would, moreover, want the feature extractor to be invariant to the digits' spatial location within the image, which leads to the requirement of translation invariance. In addition, it is desirable that the feature extractor be robust with respect to (w.r.t.) handwriting styles. This can be accomplished by demanding limited sensitivity of the features to certain non-linear deformations of the signals to be classified.

Feature extractors based on DCNNs have been applied with tremendous success in a wide range of practical machine learning tasks (Rumelhart et al., 1986; LeCun et al., 1990, 1998, 2015; Krizhevsky et al., 2012; Bengio et al., 2013; He et al., 2015; Mnih et al., 2015; Goodfellow et al., 2016; Silver et al., 2016). These networks are composed of multiple layers, each of which computes convolutional transforms, followed by the application of non-linearities and pooling operators.

The mathematical analysis of feature extractors generated by DCNNs was pioneered by Mallat in (Mallat, 2012). Mallat's theory applies to so-called scattering networks, where signals are propagated through layers that compute a semi-discrete wavelet transform (i.e., convolutions with filters that are obtained from a mother wavelet

through scaling and rotation operations), followed by the modulus non-linearity, without subsequent pooling. The resulting feature extractor is shown to be translation-invariant (asymptotically in the scale parameter of the underlying wavelet transform) and stable w.r.t. certain non-linear deformations. Moreover, Mallat's scattering networks lead to state-of-the-art results in various classification tasks (Bruna and Mallat, 2013; Andén and Mallat, 2014; Sifre, 2014).

DCNN-based feature extractors that were found to work well in practice employ a wide range of i) filters, namely pre-specified structured filters such as wavelets (Serre et al., 2005; Mutch and Lowe, 2006; Pinto et al., 2008; Jarrett et al., 2009), pre-specified unstructured filters such as random filters (Ranzato et al., 2007; Jarrett et al., 2009), and filters that are learned in a supervised (Huang and LeCun, 2006; Jarrett et al., 2009) or an unsupervised (Ranzato et al., 2006, 2007; Jarrett et al., 2009) fashion, ii) non-linearities, beyond the modulus function (Mutch and Lowe, 2006; Jarrett et al., 2009; Mallat, 2012), namely hyperbolic tangents (Huang and LeCun, 2006; Ranzato et al., 2007; Jarrett et al., 2009), rectified linear units (Nair and Hinton, 2010; Glorot et al., 2011), and logistic sigmoids (Glorot and Bengio, 2010; Mohamed et al., 2011), and iii) pooling operators, namely sub-sampling (Pinto et al., 2008), average pooling (Huang and LeCun, 2006; Jarrett et al., 2009), and max-pooling (Serre et al., 2005; Mutch and Lowe, 2006; Ranzato et al., 2007; Jarrett et al., 2009). In addition, the filters, non-linearities, and pooling operators can be different in different network layers. This motivates to develop *generalized* scattering networks that encompass all these elements in full generality, which is the first main contribution of Chapter 3.

Convolutional transforms as applied in DCNNs can be interpreted as semi-discrete signal transforms (Mallat and Zhong, 1992; Unser, 1995; Vandergheynst, 2002a; Candès and Donoho, 2005; Mallat, 2009; Grohs, 2012; Kutyniok and Labate, 2012b; Grohs et al., 2015) (i.e., convolutional transforms with filters that are countably parametrized). Corresponding prominent representatives are curvelet (Candès and Donoho, 2004, 2005; Grohs et al., 2015) and shearlet (Guo et al., 2006; Kutyniok and Labate, 2012b) transforms, both of which are known to

be highly effective in extracting features characterized by curved edges in images. The theory developed in Chapter 3 allows for general semi-discrete signal transforms, general Lipschitz-continuous non-linearities (e.g., rectified linear units, shifted logistic sigmoids, hyperbolic tangents, and modulus functions), and incorporates continuous-time Lipschitz pooling operators that emulate discrete-time sub-sampling and averaging. Finally, different network layers may be equipped with different convolutional transforms, different Lipschitz-continuous non-linearities, and different Lipschitz-continuous pooling operators.

Regarding translation invariance, it was argued, e.g., in (Serre et al., 2005; Huang and LeCun, 2006; Mutch and Lowe, 2006; Ranzato et al., 2007; Jarrett et al., 2009), that in practice invariance of the extracted features is crucially governed by the network depth and by the presence of pooling operators (such as, e.g., sub-sampling (Pinto et al., 2008), average-pooling (Huang and LeCun, 2006; Jarrett et al., 2009), or max-pooling (Serre et al., 2005; Mutch and Lowe, 2006; Ranzato et al., 2007; Jarrett et al., 2009)). We show that the generalized scattering networks considered in this thesis, indeed, exhibit such a *vertical* translation invariance and that pooling plays a crucial role in achieving it. Specifically, we prove that the depth of the network determines the extent to which the extracted features are translation-invariant. We also show that pooling is necessary to obtain vertical translation invariance as otherwise the features remain fully translation-covariant irrespective of network depth. We furthermore establish a deformation sensitivity bound valid for signal classes such as, e.g., band-limited functions, cartoon functions (Donoho, 2001) (which provide a good model for natural images such as those in the Caltech-256 (Griffin et al., 2007), CIFAR-100 (Krizhevsky, 2009), and MNIST (LeCun and Cortes, 1998) data sets), and Lipschitz functions. This bound shows that small non-linear deformations of the input signal lead to small changes in the corresponding feature vector.

In terms of mathematical techniques, we draw heavily from continuous frame theory (Ali et al., 1993; Kaiser, 1994). We develop a proof machinery that is completely detached from the structures of the semi-discrete transforms and the specific form of the Lipschitz

non-linearities and Lipschitz pooling operators. The proof of our deformation sensitivity bound is based on two key elements, namely a Lipschitz continuity property for the feature extractor and a deformation sensitivity bound for the signal class under consideration (e.g., band-limited functions, cartoon functions, and Lipschitz functions). This "decoupling" approach has important practical ramifications as it shows that whenever we have deformation sensitivity bounds for a signal class, we automatically get deformation sensitivity bounds for the DCNN feature extractor operating on that signal class. Our results hence establish that vertical translation invariance and limited sensitivity to deformations—for signal classes with inherent deformation insensitivity—are guaranteed by the network structure per se rather than the specific convolution kernels, non-linearities, and pooling operators.

1.2. ENERGY PROPAGATION IN DEEP CONVOLUTIONAL NEURAL NETWORKS (CHAPTER 4)

Many practical machine learning tasks, such as, e.g., the classification of images in the ImageNet data set, employ *very* deep networks with potentially hundreds of layers (He et al., 2015). Such network depths entail formidable computational challenges in the training phase due to the large number of parameters to be learned (e.g., in (Simonyan and Zisserman, 2014), the DCNN has 144 million parameters), and in operating the network due to the large number of convolutions that need to be carried out (e.g., the DCNN in (He et al., 2015) entails 11.3 billion FLOPS to pass a single image through the network). It is therefore paramount to understand how fast the energy contained in the signals generated in the individual network layers (a.k.a. feature maps) decays across layers. In addition, it is important that the feature vector—obtained by aggregating filtered versions of the feature maps—be informative in the sense of the only signal mapping to the

all-zeros feature vector being the zero input signal. This "trivial null-set" property for the feature extractor can be obtained by asking for the energy in the feature vector being proportional to that of the corresponding input signal, a property we shall refer to as "energy conservation".

First steps towards addressing these questions were made—for scattering network-based feature extractors—in (Waldspurger, 2015, Section 5) and (Czaja and Li, 2017). Specifically, it was shown that the energy in the feature maps generated by scattering networks employing, in every network layer, the same set of certain Parseval wavelets (Waldspurger, 2015, Section 5) or "uniform covering" (Czaja and Li, 2017) filters (both satisfying analyticity and vanishing moments conditions), the modulus non-linearity, and no pooling, decays at least exponentially fast and "strict" energy conservation (which, in turn, implies a trivial null-set) for the infinite-depth feature vector holds. Specifically, the feature map energy decay was shown to be at least of order $\mathcal{O}(a^{-N})$, for some *unspecified* $a > 1$, where N denotes the network depth. We note that d-dimensional uniform covering filters as introduced in (Czaja and Li, 2017) are a family of functions whose Fourier transforms' support sets can be covered by a union of finitely many balls. This covering condition is satisfied by, e.g., Weyl-Heisenberg filters (Gröchenig, 2001) with a band-limited prototype function, but fails to hold for multi-scale filters such as wavelets (Daubechies, 1992; Mallat, 2009), (α)-curvelets (Candès and Donoho, 2004, 2005; Grohs et al., 2015), shearlets (Guo et al., 2006; Kutyniok and Labate, 2012b), or ridgelets (Candès, 1998; Candès and Donoho, 1999; Grohs, 2012), see (Czaja and Li, 2017, Remark 2.2 (b)).

The first main contribution of Chapter 4 is a characterization of the feature map energy decay rate in scattering networks employing the modulus non-linearity, no pooling, and *general* filters that constitute a frame (Daubechies, 1992; Ali et al., 1993; Kaiser, 1994; Christensen, 2003), but not necessarily a Parseval frame, and are allowed to be different in different network layers. We find that, under mild analyticity and high-pass conditions on the filters, the energy decay rate is at least polynomial in the network depth, i.e., the decay is at least of

order $\mathcal{O}(N^{-\alpha})$, and we *explicitly* specify the decay exponent $\alpha > 0$. This result encompasses, inter alia, various constructions of Weyl-Heisenberg filters, wavelets, ridgelets, (α)-curvelets, shearlets, and learned filters (of course as long as the learning algorithm imposes the analyticity and high-pass conditions we require). For broad families of wavelets and Weyl-Heisenberg filters, the guaranteed energy decay rate is shown to be exponential in the network depth, i.e., the decay is at least of order $\mathcal{O}(a^{-N})$ where an arbitrary decay factor $a > 1$ can be realized through suitable choice of the mother wavelet bandwidth or the Weyl-Heisenberg prototype function bandwidth.

Our second main contribution in Chapter 4 shows that the energy decay results above are compatible with a trivial null-set for finite- and infinite-depth networks. Specifically, this is accomplished by establishing energy proportionality between the feature vector and the underlying input signal with the proportionality constant lower- and upper-bounded by the frame bounds of the filters employed in the different layers. We show that this energy conservation result is a consequence of a demodulation effect induced by the modulus nonlinearity in combination with the analyticity and high-pass properties of the filters. Specifically, in every network layer, the modulus nonlinearity moves the spectral content of each individual feature map to base-band (i.e., to low frequencies), where it is subsequently extracted (i.e., fed into the feature vector) by a low-pass output-generating filter.

For input signals that belong to the class of Sobolev functions[1], our energy decay and conservation results are shown to yield handy estimates of the number of layers needed to have at least $((1-\varepsilon)\cdot 100)\%$ of the input signal energy be contained in the feature vector. Finally, we show how networks of fixed (possibly small) depth N, say $N = 2$, can be designed that capture most of the input signal's energy.

We emphasize that throughout energy decay results pertain to the

[1] A wide range of practically relevant signal classes are Sobolev functions, for example, band-limited functions and—as established in the present thesis—cartoon functions (Donoho, 2001) which are a good model for natural images such as, e.g., images of handwritten digits (LeCun and Cortes, 1998).

feature maps, whereas energy conservation statements apply to the feature vector, obtained by aggregating filtered versions of the feature maps.

1.3. FROM THEORY TO PRACTICE: DISCRETE-TIME DEEP CONVOLUTIONAL NEURAL NETWORKS (CHAPTER 5)

The purpose of Chapter 5 is to build the bridge between theory and practice. Specifically, we introduce new discrete-time DCNN architectures and propose a mathematical framework for their analysis. The architectures we present incorporate general filters, Lipschitz non-linearities, and Lipschitz pooling operators, and build the feature vector from subsets of the layers. This leads us to the notions of local and global feature vector properties with globality pertaining to characteristics brought out by the union of features across all network layers, and locality identifying attributes made explicit in individual layers.

Besides providing analytical performance results of general validity, we also investigate how certain structural properties of the input signal are reflected in the corresponding feature vectors. Specifically, we analyze the (local and global) deformation and translation sensitivity properties of feature vectors corresponding to sampled cartoon functions (Donoho, 2001).

Our theoretical results are complemented by extensive numerical studies on facial landmark detection and handwritten digit classification. Specifically, we elucidate the role of local feature vector properties through a feature relevance study.

1.4. PUBLICATIONS

The majority of the results in this thesis have been published during the course of the PhD studies. Specifically, the results in Chapter 3

appear in (Wiatowski and Bölcskei, 2015, 2018; Grohs et al., 2016). Moreover, the results presented in Chapter 4 have been published in (Grohs et al., 2017; Wiatowski et al., 2017, 2018), and the results in Chapter 5 were presented in (Wiatowski et al., 2016).

CHAPTER 2

Mathematical Prerequisites

The basic building block of a DCNN consists of a convolutional transform followed by a non-linearity and a pooling operation. In this chapter, we review the theory of convolutional transforms (specifically, of semi-discrete frames) and give a list of structured example transforms of interest in the context of this thesis. Moreover, we give a brief overview of non-linearities and pooling operators that are widely used in the deep learning literature, and establish that these non-linearities and pooling operators all satisfy the Lipschitz property.

We start this chapter by introducing the notation employed in this thesis.

2.1. NOTATION

Throughout the thesis, we employ the following notation.

Scalars, vectors, matrices, and tensors

The complex conjugate of $z \in \mathbb{C}$ is denoted by $\overline{z}$. We write $\mathrm{Re}(z)$ for the real, and $\mathrm{Im}(z)$ for the imaginary part of $z \in \mathbb{C}$. The Euclidean inner product of $x, y \in \mathbb{C}^d$ is $\langle x, y \rangle := \sum_{i=1}^{d} x_i \overline{y_i}$, with associated norm $|x| := \sqrt{\langle x, x \rangle}$. We denote the identity matrix by $E \in \mathbb{R}^{d \times d}$. For the matrix $M \in \mathbb{R}^{d \times d}$, $M_{i,j}$ designates the entry in its i-th row

and j-th column, and for a tensor $T \in \mathbb{R}^{d\times d\times d}$, $T_{i,j,k}$ refers to its (i,j,k)-th component. The supremum norm of the matrix $M \in \mathbb{R}^{d\times d}$ is defined as $|M|_\infty := \sup_{i,j} |M_{i,j}|$, and the supremum norm of the tensor $T \in \mathbb{R}^{d\times d\times d}$ is $|T|_\infty := \sup_{i,j,k} |T_{i,j,k}|$.

Sets and groups

We write $B_r(x) \subseteq \mathbb{R}^d$ for the open ball of radius $r > 0$ centered at $x \in \mathbb{R}^d$. The Minkowski sum of sets $A, B \subseteq \mathbb{R}^d$ is $(A+B) := \{a+b \,|\, a \in A,\ b \in B\}$, and $A\Delta B := (A\backslash B) \cup (B\backslash A)$ denotes their symmetric difference. The cardinality of the set A is denoted by $\mathrm{card}(A)$. The indicator function of a set $B \subseteq \mathbb{R}^d$ is defined as $\mathbb{1}_B(x) = 1$, for $x \in B$, and $\mathbb{1}_B(x) = 0$, for $x \in \mathbb{R}^d\backslash B$. The support $\mathrm{supp}(f)$ of a function $f : \mathbb{R}^d \to \mathbb{C}$ is the closure of the set $\{x \in \mathbb{R}^d \mid f(x) \neq 0\}$ in the topology induced by the Euclidean norm $|\cdot|$. $O(d)$ stands for the orthogonal group of dimension $d \in \mathbb{N}$, and $SO(d)$ for the special orthogonal group. The first canonical orthant is $H := \{x \in \mathbb{R}^d \mid x_k \geqslant 0,\ k = 1,\ldots,d\}$, and we define the rotated orthant $H_A := \{Ax \mid x \in H\}$, for $A \in O(d)$.

Lebesgue-measurable functions

For a Lebesgue-measurable function $f : \mathbb{R}^d \to \mathbb{C}$, we write $\int_{\mathbb{R}^d} f(x)\mathrm{d}x$ for the integral of f w.r.t. Lebesgue measure μ_L. For $p \in [1,\infty)$, $L^p(\mathbb{R}^d)$ stands for the space of Lebesgue-measurable functions $f : \mathbb{R}^d \to \mathbb{C}$ satisfying $\|f\|_p := (\int_{\mathbb{R}^d} |f(x)|^p \mathrm{d}x)^{1/p} < \infty$. $L^\infty(\mathbb{R}^d)$ denotes the space of Lebesgue-measurable functions $f : \mathbb{R}^d \to \mathbb{C}$ such that $\|f\|_\infty := \inf\{\alpha > 0 \mid |f(x)| \leqslant \alpha \text{ for a.e.}^1\ x \in \mathbb{R}^d\} < \infty$. For $f, g \in L^2(\mathbb{R}^d)$ we set $\langle f, g\rangle := \int_{\mathbb{R}^d} f(x)\overline{g(x)}\mathrm{d}x$. For a countable set $\mathcal{Q}$, $(L^2(\mathbb{R}^d))^{\mathcal{Q}}$ denotes the space of sets $s := \{s_q\}_{q\in\mathcal{Q}}$, $s_q \in L^2(\mathbb{R}^d)$, for all $q \in \mathcal{Q}$, satisfying $|||s||| := (\sum_{q\in\mathcal{Q}} \|s_q\|_2^2)^{1/2} < \infty$. For a measurable set $B \subseteq \mathbb{R}^d$, we let $\mathrm{vol}^d(B) := \int_{\mathbb{R}^d} \mathbb{1}_B(x)\mathrm{d}x = \int_B 1\mathrm{d}x$.

[1] Throughout "a.e." is w.r.t. Lebesgue measure.

Linear operators

Id : $L^p(\mathbb{R}^d) \to L^p(\mathbb{R}^d)$ stands for the identity operator on $L^p(\mathbb{R}^d)$. We denote the Fourier transform of $f \in L^1(\mathbb{R}^d)$ by $\widehat{f}(\omega) := \int_{\mathbb{R}^d} f(x)e^{-2\pi i\langle x,\omega\rangle}\mathrm{d}x$ and extend it in the usual way to $L^2(\mathbb{R}^d)$ (Rudin, 1991, Theorem 7.9). The convolution of $f \in L^2(\mathbb{R}^d)$ and $g \in L^1(\mathbb{R}^d)$ is $(f * g)(y) := \int_{\mathbb{R}^d} f(x)g(y-x)\mathrm{d}x$. We write $(T_t f)(x) := f(x-t)$, $t \in \mathbb{R}^d$, for the translation operator, and $(M_\omega f)(x) := e^{2\pi i\langle x,\omega\rangle} f(x)$, $\omega \in \mathbb{R}^d$, for the modulation operator. Involution is defined by $(If)(x) := \overline{f(-x)}$. The operator norm of the bounded linear operator $A : L^p(\mathbb{R}^d) \to L^q(\mathbb{R}^d)$ is $\|A\|_{p,q} := \sup_{\|f\|_p=1} \|Af\|_q$.

Differentiable functions and vector fields

$H^s(\mathbb{R}^d)$, with $s > 0$, stands for the Sobolev space of functions $f \in L^2(\mathbb{R}^d)$ satisfying $\|f\|_{H^s} := (\int_{\mathbb{R}^d} |\widehat{f}(\omega)|^2(1+|\omega|^2)^s\mathrm{d}\omega)^{1/2} < \infty$, see (Grafakos, 2009, Section 6.2.1). Here, the index s reflects the "degree" of smoothness of $f \in H^s(\mathbb{R}^d)$, i.e., larger s entails smoother f. For a multi-index $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{N}_0^d$, D^α denotes the differential operator $D^\alpha := (\partial/\partial x_1)^{\alpha_1} \ldots (\partial/\partial x_d)^{\alpha_d}$, with order $|\alpha| := \sum_{i=1}^d \alpha_i$. If $|\alpha| = 0$, $D^\alpha f := f$, for $f : \mathbb{R}^d \to \mathbb{C}$. The space of functions $f : \mathbb{R}^d \to \mathbb{C}$ whose derivatives $D^\alpha f$ of order at most $N \in \mathbb{N}_0$ are continuous is designated by $C^N(\mathbb{R}^d, \mathbb{C})$, and the space of infinitely differentiable functions is $C^\infty(\mathbb{R}^d, \mathbb{C})$. $S(\mathbb{R}^d, \mathbb{C})$ stands for the Schwartz space, i.e., the space of functions $f \in C^\infty(\mathbb{R}^d, \mathbb{C})$ whose derivatives $D^\alpha f$ along with the function itself are rapidly decaying (Rudin, 1991, Section 7.3) in the sense of $\sup_{|\alpha|\leqslant N} \sup_{x\in\mathbb{R}^d}(1+|x|^2)^N|(D^\alpha f)(x)| < \infty$, for all $N \in \mathbb{N}_0$. We denote the gradient of a function $f : \mathbb{R}^d \to \mathbb{C}$ as ∇f. The space of continuous vector fields $v : \mathbb{R}^p \to \mathbb{R}^q$ is $C(\mathbb{R}^p, \mathbb{R}^q)$, and for $k, p, q \in \mathbb{N}$, the space of k-times continuously differentiable vector fields $v : \mathbb{R}^p \to \mathbb{R}^q$ is written as $C^k(\mathbb{R}^p, \mathbb{R}^q)$. For a vector field $v : \mathbb{R}^d \to \mathbb{R}^d$, we let Dv be its Jacobian matrix, and $D^2 v$ its Jacobian tensor, with associated norms $\|v\|_\infty := \sup_{x\in\mathbb{R}^d} |v(x)|$, $\|Dv\|_\infty := \sup_{x\in\mathbb{R}^d} |(Dv)(x)|_\infty$, and $\|D^2 v\|_\infty := \sup_{x\in\mathbb{R}^d} |(D^2 v)(x)|_\infty$.

Miscellaneous

For $x \in \mathbb{R}$, we set $(x)_+ := \max\{0, x\}$ and $\langle x \rangle := (1 + |x|^2)^{1/2}$. The tensor product of functions $f, g : \mathbb{R}^d \to \mathbb{C}$ is $(f \otimes g)(x, y) := f(x)g(y)$, $(x, y) \in \mathbb{R}^d \times \mathbb{R}^d$. For functions $W : \mathbb{N} \to \mathbb{R}$ and $G : \mathbb{N} \to \mathbb{R}$, we say that $W(N) = \mathcal{O}(G(N))$ if there exist $C > 0$ and $N_0 \in \mathbb{N}$ such that $W(N) \leqslant CG(N)$, for all $N \geqslant N_0$.

2.2. CONVOLUTIONAL TRANSFORMS: SEMI-DISCRETE FRAMES

This section gives a brief review of the theory of semi-discrete frames which are instances of *continuous* frames (Ali et al., 1993; Kaiser, 1994), and appear in the mathematical signal processing literature, e.g., in the context of translation-covariant signal decompositions (Mallat and Zhong, 1992; Unser, 1995; Vandergheynst, 2002a), and as an intermediate step in the construction of various *fully-discrete* frames (Candès and Donoho, 2005; Grohs, 2012; Kutyniok and Labate, 2012a; Grohs et al., 2015). A list of structured example frames of interest in the context of this thesis is provided in Section 2.2.1 for the 1-D case, and in Section 2.2.2 for the 2-D case.

We first collect some basic results on semi-discrete frames.

Definition 1. *Let $\{g_\lambda\}_{\lambda \in \Lambda} \subseteq L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$ be a set of functions indexed by a countable set Λ. The collection*

$$\Psi_\Lambda := \{T_b I g_\lambda\}_{(\lambda, b) \in \Lambda \times \mathbb{R}^d}$$

is a semi-discrete frame for $L^2(\mathbb{R}^d)$ if there exist constants $A, B > 0$ such that

$$A\|f\|_2^2 \leqslant \sum_{\lambda \in \Lambda} \int_{\mathbb{R}^d} |\langle f, T_b I g_\lambda \rangle|^2 \mathrm{d}b = \sum_{\lambda \in \Lambda} \|f * g_\lambda\|_2^2 \leqslant B\|f\|_2^2, \quad (2.1)$$

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

The frame operator associated with the semi-discrete frame Ψ_Λ is defined in the weak sense as $S_\Lambda : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$,

$$S_\Lambda f := \sum_{\lambda\in\Lambda} \int_{\mathbb{R}^d} \langle f, T_b I g_\lambda\rangle (T_b I g_\lambda)\, \mathrm{d}b = \Big(\sum_{\lambda\in\Lambda} g_\lambda * I g_\lambda \Big) * f, \qquad (2.2)$$

where $\langle f, T_b I g_\lambda\rangle = (f * g_\lambda)(b)$, $(\lambda, b) \in \Lambda \times \mathbb{R}^d$, are called the frame coefficients. S_Λ is a bounded, positive, and boundedly invertible operator (Ali et al., 1993).

The reader might want to think of semi-discrete frames as shift-invariant frames (Ron and Shen, 1995; Janssen, 1998) with a continuous translation parameter $b \in \mathbb{R}$, and of the countable index set Λ as labeling a collection of scales, directions, or frequency-shifts, hence the terminology *semi-discrete*.

The following result gives a so-called Littlewood-Paley condition (Frazier et al., 1991; Daubechies, 1992) for the collection $\Psi_\Lambda = \{T_b I g_\lambda\}_{(\lambda,b)\in\Lambda\times\mathbb{R}^d}$ to form a semi-discrete frame.

Proposition 1. *(Mallat, 2009, Theorem 5.11) Let Λ be a countable set. The collection $\Psi_\Lambda = \{T_b I g_\lambda\}_{(\lambda,b)\in\Lambda\times\mathbb{R}^d}$ with atoms $\{g_\lambda\}_{\lambda\in\Lambda} \subseteq L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$ is a semi-discrete frame for $L^2(\mathbb{R}^d)$ with frame bounds $A, B > 0$ if and only if*

$$A \leqslant \sum_{\lambda\in\Lambda} |\widehat{g_\lambda}(\omega)|^2 \leqslant B, \qquad \text{a.e. } \omega \in \mathbb{R}^d. \qquad (2.3)$$

Remark 1. *What is behind Proposition 1 is a result on the unitary equivalence between operators (Naylor and Sell, 1982, Definition 5.19.3). Specifically, Proposition 1 follows from the fact that the multiplier $\sum_{\lambda\in\Lambda} |\widehat{g_\lambda}|^2$ is unitarily equivalent to the frame operator S_Λ in* (2.2) *according to*

$$\mathcal{F} S_\Lambda \mathcal{F}^{-1} = \sum_{\lambda\in\Lambda} |\widehat{g_\lambda}|^2,$$

where $\mathcal{F} : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ denotes the Fourier transform. We refer the interested reader to (Bölcskei et al., 1998), where the framework of unitary equivalence was formalized in the context of shift-invariant frames for $\ell^2(\mathbb{Z})$.

The following proposition states normalization results for semi-discrete frames that come in handy in satisfying, e.g., the admissibility condition (3.14) as discussed in Section 3.2, or the condition (4.20) on the product of the frame lower and frame upper bounds as discussed in Section 4.3.

Proposition 2. *Let $\Psi_\Lambda = \{T_b I g_\lambda\}_{(\lambda,b)\in\Lambda\times\mathbb{R}^d}$ be a semi-discrete frame for $L^2(\mathbb{R}^d)$ with frame bounds A, B.*

i) For $C > 0$, the family of functions $\widetilde{\Psi}_\Lambda := \{T_b I\, \widetilde{g_\lambda}\}_{(\lambda,b)\in\Lambda\times\mathbb{R}^d}$,

$$\widetilde{g_\lambda} := C^{-1/2} g_\lambda, \quad \forall \lambda \in \Lambda,$$

is a semi-discrete frame for $L^2(\mathbb{R}^d)$ with frame bounds $\widetilde{A} := \frac{A}{C}$ and $\widetilde{B} := \frac{B}{C}$.

ii) The family of functions $\Psi_\Lambda^\natural := \{T_b I g_\lambda^\natural\}_{(\lambda,b)\in\Lambda\times\mathbb{R}^d}$,

$$g_\lambda^\natural := \mathcal{F}^{-1}\Big(\widehat{g_\lambda}\Big(\sum_{\lambda'\in\Lambda} |\widehat{g_{\lambda'}}|^2\Big)^{-1/2}\Big), \quad \forall \lambda \in \Lambda,$$

is a semi-discrete Parseval frame for $L^2(\mathbb{R}^d)$, i.e., the frame bounds satisfy $A^\natural = B^\natural = 1$.

Proof. We start by proving statement i). As Ψ_Λ is a frame for $L^2(\mathbb{R}^d)$, we have

$$A\|f\|_2^2 \leqslant \sum_{\lambda\in\Lambda} \|f * g_\lambda\|_2^2 \leqslant B\|f\|_2^2, \qquad \forall f \in L^2(\mathbb{R}^d). \tag{2.4}$$

With $g_\lambda = \sqrt{C}\widetilde{g_\lambda}$, for all $\lambda \in \Lambda$, in (2.4) we get $A\|f\|_2^2 \leqslant \sum_{\lambda\in\Lambda} \|f * \sqrt{C}\widetilde{g_\lambda}\|_2^2 \leqslant B\|f\|_2^2$, for all $f \in L^2(\mathbb{R}^d)$, which is equivalent to $\frac{A}{C}\|f\|_2^2 \leqslant \sum_{\lambda\in\Lambda} \|f * \widetilde{g_\lambda}\|_2^2 \leqslant \frac{B}{C}\|f\|_2^2$, for all $f \in L^2(\mathbb{R}^d)$, and hence establishes i). To prove statement ii), we first note that $\mathcal{F}g_\lambda^\natural = \widehat{g_\lambda}\big(\sum_{\lambda'\in\Lambda} |\widehat{g_{\lambda'}}|^2\big)^{-1/2}$, for all $\lambda \in \Lambda$, and thus $\sum_{\lambda\in\Lambda} |(\mathcal{F}g_\lambda^\natural)(\omega)|^2 = \sum_{\lambda\in\Lambda} |\widehat{g_\lambda}(\omega)|^2 \Big(\sum_{\lambda'\in\Lambda} |\widehat{g_{\lambda'}}(\omega)|^2\Big)^{-1} = 1$, a.e. $\omega \in \mathbb{R}^d$. Application of Proposition 1 then establishes that $\Psi_\Lambda^\natural$ is a semi-discrete Parseval frame for $L^2(\mathbb{R}^d)$, i.e., the frame bounds satisfy $A^\natural = B^\natural = 1$. □

2.2.1. Examples of semi-discrete frames in 1-D

General 1-D semi-discrete frames are given by collections

$$\Psi = \{T_b I g_k\}_{(k,b)\in\mathbb{Z}\times\mathbb{R}} \tag{2.5}$$

with atoms $g_k \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$, indexed by the integers $\Lambda = \mathbb{Z}$, and satisfying the Littlewood-Paley condition

$$A \leqslant \sum_{k\in\mathbb{Z}} |\widehat{g_k}(\omega)|^2 \leqslant B, \qquad a.e.\ \omega \in \mathbb{R}. \tag{2.6}$$

The structural example frames we consider in this section are Weyl-Heisenberg (Gabor) frames (where the g_k are obtained through modulation from a prototype function) and wavelet frames (where the g_k are obtained through scaling from a mother wavelet).

Semi-discrete Weyl-Heisenberg frames

Weyl-Heisenberg frames (Daubechies et al., 1986, 1995; Janssen, 1995; Gröchenig, 2001) (a.k.a. Gabor frames) are well-suited to the extraction of sinusoidal features (Gröchenig and Samarah, 2000), and have been applied successfully in various practical feature extraction tasks (Lee et al., 2009; Ellis et al., 2011). A semi-discrete Weyl-Heisenberg frame for $L^2(\mathbb{R})$ is a collection of functions according to (2.5), where $g_k(x) := e^{2\pi i k x} g(x)$, $k \in \mathbb{Z}$, with the prototype function $g \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$. The atoms $\{g_k\}_{k\in\mathbb{Z}}$ satisfy the Littlewood-Paley condition (2.6) according to

$$A \leqslant \sum_{k\in\mathbb{Z}} |\widehat{g}(\omega - k)|^2 \leqslant B, \qquad a.e.\ \omega \in \mathbb{R}. \tag{2.7}$$

A popular function $g \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ satisfying (2.7) is the Gaussian function (Gröchenig, 2001).

Semi-discrete wavelet frames

Wavelets are well-suited to the extraction of signal features characterized by singularities (Daubechies, 1992; Mallat and Zhong, 1992), and

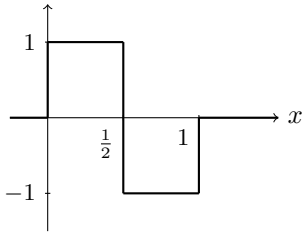

Fig. 2.1: The Haar wavelet $\psi(x)$ in 1-D.

have been applied successfully in various practical feature extraction tasks (Lin and Qu, 2000; Tzanetakis and Cook, 2002). A semi-discrete wavelet frame for $L^2(\mathbb{R})$ is a collection of functions according to (2.5), where $g_k(x) := 2^k\psi(2^k x)$, $k \in \mathbb{Z}$, with the mother wavelet $\psi \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$. The atoms $\{g_k\}_{k\in\mathbb{Z}}$ satisfy the Littlewood-Paley condition (2.6) according to

$$A \leqslant \sum_{k\in\mathbb{Z}} |\widehat{\psi}(2^{-k}\omega)|^2 \leqslant B, \qquad a.e.\ \omega \in \mathbb{R}. \tag{2.8}$$

A large class of functions ψ satisfying (2.8) can be obtained through a multi-resolution analysis in $L^2(\mathbb{R})$ (Mallat, 2009, Definition 7.1) such as, e.g., the Haar wavelet (see Fig. 2.1).

2.2.2. Examples of semi-discrete frames in 2-D

Semi-discrete wavelet frames

Two-dimensional wavelets are well-suited to the extraction of signal features characterized by point singularities (such as, e.g., stars in astronomical images (Kutyniok and Donoho, 2013)), and have been applied successfully in various practical feature extraction tasks, e.g., in (Unser, 1995; Serre et al., 2005; Mutch and Lowe, 2006; Pinto et al., 2008). Prominent families of two-dimensional wavelet frames are tensor wavelet frames and directional wavelet frames.

Tensor wavelets

A semi-discrete tensor wavelet frame for $L^2(\mathbb{R}^2)$ is a collection of functions according to

$$\Psi_{\Lambda_{\mathrm{TW}}} := \{T_b I g_{(e,j)}\}_{(e,j)\in\Lambda_{\mathrm{TW}}, b\in\mathbb{R}^2}, \qquad g_{(e,j)}(x) := 2^{2j}\psi^e(2^j x),$$

where

$$\Lambda_{\mathrm{TW}} := \big\{((0,0),0)\big\} \cup \big\{(e,j) \mid e \in E\backslash\{(0,0)\},\ j \geqslant 0\big\},$$

and $E := \{0,1\}^2$. Here, the functions $\psi^e \in L^1(\mathbb{R}^2)\cap L^2(\mathbb{R}^2)$ are tensor products of a coarse-scale function $\phi \in L^1(\mathbb{R})\cap L^2(\mathbb{R})$ and a fine-scale function $\psi \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ according to

$$\psi^{(0,0)} := \phi\otimes\phi, \qquad \psi^{(1,0)} := \psi\otimes\phi, \qquad \psi^{(0,1)} := \phi\otimes\psi, \qquad \psi^{(1,1)} := \psi\otimes\psi.$$

The corresponding Littlewood-Paley condition (2.3) reads

$$A \leqslant \big|\widehat{\psi^{(0,0)}}(\omega)\big|^2 + \sum_{j\geqslant 0}\sum_{e\in E\backslash\{(0,0)\}} |\widehat{\psi^e}(2^{-j}\omega)|^2 \leqslant B, \tag{2.9}$$

for a.e. $\omega \in \mathbb{R}^2$. A large class of functions ϕ, ψ satisfying (2.9) can be obtained through a multi-resolution analysis in $L^2(\mathbb{R})$ (Mallat, 2009, Definition 7.1).

Directional wavelets

A semi-discrete directional wavelet frame for $L^2(\mathbb{R}^2)$ is a collection of functions according to

$$\Psi_{\Lambda_{\mathrm{DW}}} := \{T_b I g_{(j,k)}\}_{(j,k)\in\Lambda_{\mathrm{DW}}, b\in\mathbb{R}^2},$$

with

$$g_{(-J,0)}(x) := 2^{-2J}\phi(2^{-J}x), \qquad g_{(j,k)}(x) := 2^{2j}\psi(2^j R_{\theta_k} x),$$

where $\Lambda_{\mathrm{DW}} := \big\{(-J,0)\big\} \cup \big\{(j,k) \mid j \in \mathbb{Z} \text{ with } j > -J,\ k \in \{0,\dots,K-1\}\big\}$, R_θ is a 2×2 rotation matrix defined as

$$R_\theta := \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}, \qquad \theta \in [0, 2\pi), \tag{2.10}$$

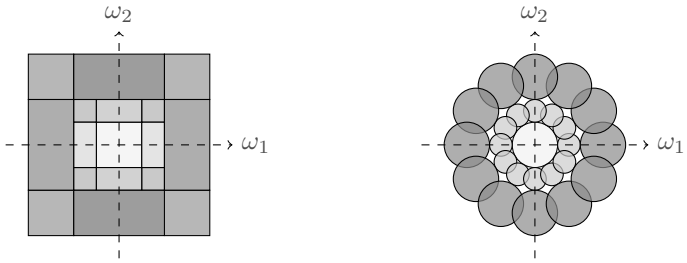

Fig. 2.2: Partitioning of the frequency plane $\mathbb{R}^2$ induced by (left) a semi-discrete tensor wavelet frame, and (right) a semi-discrete directional wavelet frame.

and $\theta_k := \frac{2\pi k}{K}$, with $k = 0, \ldots, K-1$, for a fixed $K \in \mathbb{N}$, are rotation angles. The functions $\phi \in L^1(\mathbb{R}^2) \cap L^2(\mathbb{R}^2)$ and $\psi \in L^1(\mathbb{R}^2) \cap L^2(\mathbb{R}^2)$ are referred to in the literature as coarse-scale wavelet and fine-scale wavelet, respectively. The integer $J \in \mathbb{Z}$ corresponds to the coarsest scale resolved and the atoms $\{g_{(j,k)}\}_{(j,k)\in\Lambda_{\mathrm{DW}}}$ satisfy the Littlewood-Paley condition (2.3) according to

$$A \leqslant |\widehat{\phi}(2^J\omega)|^2 + \sum_{j>-J} \sum_{k=0}^{K-1} |\widehat{\psi}(2^{-j}R_{\theta_k}\omega)|^2 \leqslant B, \tag{2.11}$$

for a.e. $\omega \in \mathbb{R}^2$. Prominent examples of functions ϕ, ψ satisfying (2.11) are the Gaussian function for ϕ and a modulated Gaussian function for ψ (Mallat, 2009).

Semi-discrete ridgelet frames

Ridgelets, introduced in (Candès, 1998; Candès and Donoho, 1999), are well-suited to the extraction of signal features characterized by straight-line singularities (such as, e.g., straight edges in images), and have been applied successfully in various practical feature extraction tasks (Chen et al., 2005; Arivazhagan et al., 2006; Dettori and Semler, 2007; Qiao et al., 2010).

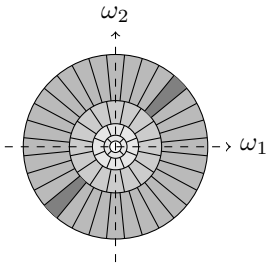

Fig. 2.3: Partitioning of the frequency plane $\mathbb{R}^2$ induced by a semi-discrete ridgelet frame.

A semi-discrete ridgelet frame for $L^2(\mathbb{R}^2)$ is a collection of functions according to

$$\Psi_{\Lambda_{\mathrm{R}}} := \{T_b I g_{(j,l)}\}_{(j,l)\in\Lambda_{\mathrm{R}}, b\in\mathbb{R}^2},$$

with

$$g_{(0,0)}(x) := \phi(x), \qquad g_{(j,l)}(x) := \psi_{(j,l)}(x),$$

where $\Lambda_{\mathrm{R}} := \big\{(0,0)\big\} \cup \big\{(j,l) \mid j \geqslant 1,\ l = 1,\dots,2^j - 1\big\}$, and the atoms $\{g_{(j,l)}\}_{(j,l)\in\Lambda_{\mathrm{R}}}$ satisfy the Littlewood-Paley condition (2.3) according to

$$A \leqslant |\widehat{\phi}(\omega)|^2 + \sum_{j=1}^{\infty}\sum_{l=1}^{2^j-1} |\widehat{\psi_{(j,l)}}(\omega)|^2 \leqslant B, \qquad \text{a.e. } \omega \in \mathbb{R}^2. \tag{2.12}$$

The functions $\psi_{(j,l)} \in L^1(\mathbb{R}^2) \cap L^2(\mathbb{R}^2)$, $(j,l) \in \Lambda_{\mathrm{R}} \backslash \{(0,0)\}$, are designed to be constant in the direction specified by the parameter l, and to have a Fourier transform $\widehat{\psi_{(j,l)}}$ supported on a pair of opposite wedges of size $2^{-j} \times 2^j$ in the dyadic corona $\{\omega \in \mathbb{R}^2 \mid 2^j \leqslant |\omega| \leqslant 2^{j+1}\}$, see Fig. 2.3. We refer the reader to (Grohs, 2012, Proposition 6) for constructions of functions $\phi, \psi_{(j,l)}$ satisfying (2.12) with $A = B = 1$.

Semi-discrete curvelet frames

Curvelets, introduced in (Candès and Donoho, 2004, 2005), are well-suited to the extraction of signal features characterized by curve-

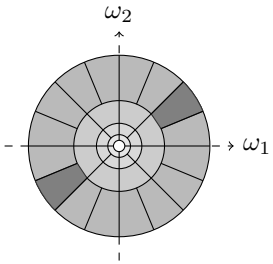

Fig. 2.4: Partitioning of the frequency plane $\mathbb{R}^2$ induced by a semi-discrete curvelet frame.

like singularities (such as, e.g., curved edges in images), and have been applied successfully in various practical feature extraction tasks (Dettori and Semler, 2007; Ma and Plonka, 2010).

A semi-discrete curvelet frame for $L^2(\mathbb{R}^2)$ is a collection of functions according to

$$\Psi_{\Lambda_C} := \{T_b I g_{(j,l)}\}_{(j,l)\in\Lambda_C, b\in\mathbb{R}^2},$$

with

$$g_{(-1,0)}(x) := \phi(x), \qquad g_{(j,l)}(x) := \psi_j(R_{\theta_{j,l}} x),$$

where $\Lambda_C := \big\{(-1,0)\big\} \cup \big\{(j,l) \mid j \geqslant 0,\ l = 0,\dots,L_j - 1\big\}$, $R_\theta \in \mathbb{R}^{2\times 2}$ is the rotation matrix defined in (2.10), and $\theta_{j,l} := \pi l 2^{-\lceil j/2\rceil - 1}$, for $j \geqslant 0$, and $0 \leqslant l < L_j := 2^{\lceil j/2\rceil + 2}$, are scale-dependent rotation angles. The functions $\phi \in L^1(\mathbb{R}^2) \cap L^2(\mathbb{R}^2)$ and $\psi_j \in L^1(\mathbb{R}^2) \cap L^2(\mathbb{R}^2)$ satisfy the Littlewood-Paley condition (2.3) according to

$$A \leqslant |\widehat{\phi}(\omega)|^2 + \sum_{j=0}^{\infty} \sum_{l=0}^{L_j - 1} |\widehat{\psi_j}(R_{\theta_{j,l}}\omega)|^2 \leqslant B, \qquad \text{a.e. } \omega \in \mathbb{R}^2. \quad (2.13)$$

The functions ψ_j, $j \geqslant 0$, are designed to have their Fourier transform $\widehat{\psi_j}$ supported on a pair of opposite wedges of size $2^{-j/2} \times 2^j$ in the dyadic corona $\{\omega \in \mathbb{R}^2 \mid 2^j \leqslant |\omega| \leqslant 2^{j+1}\}$, see Fig. 2.4. We refer the reader to (Candès and Donoho, 2005, Theorem 4.1) for constructions of functions ϕ, ψ_j satisfying (2.13) with $A = B = 1$.

Remark 2. *For further examples of interesting structured semi-discrete frames, we refer to (Kutyniok and Labate, 2012b), which discusses semi-discrete shearlet frames, and (Grohs et al., 2015), which deals with semi-discrete α-curvelet frames.*

2.3. NON-LINEARITIES

This section gives a brief overview of non-linearities $M : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ that are widely used in the deep learning literature and that fit into our theory. For each example, we establish that it satisfies the following conditions:

i) Lipschitz continuity: There exists a constant $L \geqslant 0$ such that

$$\|Mf - Mh\|_2 \leqslant L\|f - h\|_2, \qquad \forall f, h \in L^2(\mathbb{R}^d).$$

ii) $Mf = 0$ for $f = 0$.

All non-linearities considered here are pointwise (also referred to as memoryless in the mathematical signal processing literature) operators in the sense of

$$M : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d), \qquad (Mf)(x) = \rho(f(x)), \tag{2.14}$$

where $\rho : \mathbb{C} \to \mathbb{C}$. An immediate consequence of this property is that the operator M commutes with the translation operator T_t:

$$(MT_t f)(x) = \rho((T_t f)(x)) = \rho(f(x - t)) = T_t \rho(f(x)) = (T_t M f)(x),$$

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

Modulus function

The modulus function

$$|\cdot| : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d), \qquad |f|(x) := |f(x)|,$$

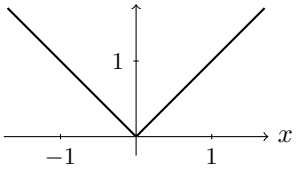

Fig. 2.5: The modulus non-linearity on $\mathbb{R}$.

has been applied successfully in the deep learning literature, e.g., in (Mutch and Lowe, 2006; Jarrett et al., 2009), and most prominently in scattering networks (Mallat, 2012). Lipschitz continuity with $L = 1$ follows from

$$\begin{aligned}\||f| - |h|\|_2^2 &= \int_{\mathbb{R}^d} ||f(x)| - |h(x)||^2 \mathrm{d}x \\ &\leqslant \int_{\mathbb{R}^d} |f(x) - h(x)|^2 \mathrm{d}x = \|f - h\|_2^2, \qquad \forall f, h \in L^2(\mathbb{R}^d),\end{aligned}$$

by the reverse triangle inequality. Furthermore, obviously $|f| = 0$ for $f = 0$, and finally $|\cdot|$ is pointwise as (2.14) is satisfied with $\rho(x) := |x|$.

Rectified linear unit

The rectified linear unit non-linearity (Nair and Hinton, 2010; Glorot et al., 2011) (a.k.a. ReLU) is defined as $R : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$,

$$(Rf)(x) := \max\{0, \mathrm{Re}(f(x))\} + i \max\{0, \mathrm{Im}(f(x))\}.$$

We start by establishing that R is Lipschitz-continuous with $L = 2$. To this end, fix $f, h \in L^2(\mathbb{R}^d)$. We have

$$\begin{aligned}&|(Rf)(x) - (Rh)(x)| = \big|\max\{0, \mathrm{Re}(f(x))\} + i \max\{0, \mathrm{Im}(f(x))\} \\ &- \big(\max\{0, \mathrm{Re}(h(x))\} + i \max\{0, \mathrm{Im}(h(x))\}\big)\big| \\ &\leqslant \big|\max\{0, \mathrm{Re}(f(x))\} - \max\{0, \mathrm{Re}(h(x))\}\big| \qquad (2.15) \\ &+ \big|\max\{0, \mathrm{Im}(f(x))\} - \max\{0, \mathrm{Im}(h(x))\}\big| \\ &\leqslant \big|\mathrm{Re}(f(x)) - \mathrm{Re}(h(x))\big| + \big|\mathrm{Im}(f(x)) - \mathrm{Im}(h(x))\big| \qquad (2.16) \\ &\leqslant \big|f(x) - h(x)\big| + \big|f(x) - h(x)\big| = 2|f(x) - h(x)|, \qquad (2.17)\end{aligned}$$

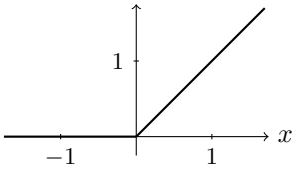

Fig. 2.6: The rectified linear unit non-linearity on $\mathbb{R}$.

where we used the triangle inequality in (2.15),

$$|\max\{0,a\} - \max\{0,b\}| \leqslant |a-b|, \qquad \forall a,b \in \mathbb{R},$$

in (2.16), and the Lipschitz continuity (with $L=1$) of $\mathrm{Re} : \mathbb{C} \to \mathbb{R}$ and $\mathrm{Im} : \mathbb{C} \to \mathbb{R}$ in (2.17). We therefore get

$$\begin{aligned}\|Rf - Rh\|_2 &= \Big(\int_{\mathbb{R}^d} |(Rf)(x) - (Rh)(x)|^2 \mathrm{d}x \Big)^{1/2} \\ &\leqslant 2 \Big(\int_{\mathbb{R}^d} |f(x) - h(x)|^2 \mathrm{d}x \Big)^{1/2} = 2\,\|f-h\|_2,\end{aligned}$$

which establishes Lipschitz continuity of R with Lipschitz constant $L=2$. Furthermore, obviously $Rf = 0$ for $f = 0$, and finally (2.14) is satisfied with $\rho(x) := \max\{0, \mathrm{Re}(x)\} + i \max\{0, \mathrm{Im}(x)\}$.

Hyperbolic tangent

The hyperbolic tangent non-linearity (see, e.g., (Huang and LeCun, 2006; Ranzato et al., 2007; Jarrett et al., 2009)) is defined as $H : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$,

$$(Hf)(x) := \tanh(\mathrm{Re}(f(x))) + i \tanh(\mathrm{Im}(f(x))),$$

where $\tanh(x) := \frac{e^x - e^{-x}}{e^x + e^{-x}}$. We start by proving that H is Lipschitz-continuous with $L=2$. To this end, fix $f, h \in L^2(\mathbb{R}^d)$. We have

$$\begin{aligned}|(Hf)(x) - (Hh)(x)| = \big| &\tanh(\mathrm{Re}(f(x))) + i \tanh(\mathrm{Im}(f(x))) \\ &- \big(\tanh(\mathrm{Re}(h(x))) + i \tanh(\mathrm{Im}(h(x))) \big) \big|\end{aligned}$$

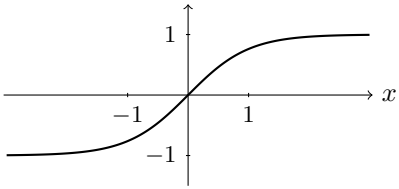

Fig. 2.7: The hyperbolic tangent non-linearity on $\mathbb{R}$.

$$\begin{aligned} &\leqslant \big|\tanh(\mathrm{Re}(f(x))) - \tanh(\mathrm{Re}(h(x)))\big| \\ &+ \big|\tanh(\mathrm{Im}(f(x))) - \tanh(\mathrm{Im}(h(x)))\big|, \quad (2.18) \end{aligned}$$

where, again, we used the triangle inequality. In order to further upper-bound (2.18), we show that tanh is Lipschitz-continuous. To this end, we make use of the following result.

Lemma 1. *(Searcóid, 2007, Theorem 9.5.1) Let $h : \mathbb{R} \to \mathbb{R}$ be a continuously differentiable function satisfying $\sup_{x\in\mathbb{R}} |h'(x)| \leqslant L$. Then, h is Lipschitz-continuous with Lipschitz constant L.*

Since $\tanh'(x) = 1 - \tanh^2(x)$, $x \in \mathbb{R}$, we have $\sup_{x\in\mathbb{R}} |\tanh'(x)| \leqslant 1$. By Lemma 1 we can therefore conclude that tanh is Lipschitz-continuous with $L = 1$, which when used in (2.18), yields

$$\begin{aligned} &|(Hf)(x) - (Hh)(x)| \\ &\leqslant \big|\mathrm{Re}(f(x)) - \mathrm{Re}(h(x))\big| + \big|\mathrm{Im}(f(x)) - \mathrm{Im}(h(x))\big| \\ &\leqslant \big|f(x) - h(x)\big| + \big|f(x) - h(x)\big| = 2|f(x) - h(x)|. \end{aligned}$$

Here, again, we used the Lipschitz continuity (with $L = 1$) of $\mathrm{Re} : \mathbb{C} \to \mathbb{R}$ and $\mathrm{Im} : \mathbb{C} \to \mathbb{R}$. Putting things together, we obtain

$$\begin{aligned} \|Hf - Hh\|_2 &= \Big(\int_{\mathbb{R}^d} |(Hf)(x) - (Hh)(x)|^2 \mathrm{d}x \Big)^{1/2} \\ &\leqslant 2 \Big(\int_{\mathbb{R}^d} |f(x) - h(x)|^2 \mathrm{d}x \Big)^{1/2} = 2\,\|f - h\|_2, \end{aligned}$$

which proves that H is Lipschitz-continuous with $L = 2$. Since $\tanh(0) = 0$, we trivially have $Hf = 0$ for $f = 0$. Finally, (2.14) is satisfied with $\rho(x) := \tanh(\mathrm{Re}(x)) + i \tanh(\mathrm{Im}(x))$.

Shifted logistic sigmoid

The shifted logistic sigmoid non-linearity[2] (see, e.g., (Glorot and Bengio, 2010; Mohamed et al., 2011)) is defined as $P : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$,

$$(Pf)(x) := \mathrm{sig}(\mathrm{Re}(f(x))) + i\,\mathrm{sig}(\mathrm{Im}(f(x))),$$

where $\mathrm{sig}(x) := \frac{1}{1+e^{-x}} - \frac{1}{2}$. We first establish that P is Lipschitz-continuous with $L = \frac{1}{2}$. To this end, fix $f, h \in L^2(\mathbb{R}^d)$. We have

$$\begin{aligned}
|(Pf)(x) - (Ph)(x)| &= \big|\,\mathrm{sig}(\mathrm{Re}(f(x))) + i\,\mathrm{sig}(\mathrm{Im}(f(x))) \\
&\quad - \big(\mathrm{sig}(\mathrm{Re}(h(x))) + i\,\mathrm{sig}(\mathrm{Im}(h(x)))\big)\big| \\
&\leqslant \big|\,\mathrm{sig}(\mathrm{Re}(f(x))) - \mathrm{sig}(\mathrm{Re}(h(x)))\big| \\
&\quad + \big|\,\mathrm{sig}(\mathrm{Im}(f(x))) - \mathrm{sig}(\mathrm{Im}(h(x)))\big|, \qquad (2.19)
\end{aligned}$$

where, again, we employed the triangle inequality. As before, to further upper-bound (2.19), we show that sig is Lipschitz-continuous. Specifically, we apply Lemma 1 with $\mathrm{sig}'(x) = \frac{e^{-x}}{(1+e^{-x})^2}$, $x \in \mathbb{R}$, and hence $\sup_{x\in\mathbb{R}} |\,\mathrm{sig}'(x)| \leqslant \frac{1}{4}$, to conclude that sig is Lipschitz-continuous with $L = \frac{1}{4}$. When used in (2.19) this yields (together with the Lipschitz continuity (with $L = 1$) of $\mathrm{Re} : \mathbb{C} \to \mathbb{R}$ and $\mathrm{Im} : \mathbb{C} \to \mathbb{R}$)

$$\begin{aligned}
&|(Pf)(x) - (Ph)(x)| \\
&\leqslant \frac{1}{4}\Big|\,\mathrm{Re}(f(x)) - \mathrm{Re}(h(x))\Big| + \frac{1}{4}\Big|\,\mathrm{Im}(f(x)) - \mathrm{Im}(h(x))\Big| \\
&\leqslant \frac{1}{4}\Big|f(x) - h(x)\Big| + \frac{1}{4}\Big|f(x) - h(x)\Big|
\end{aligned}$$

[2] Strictly speaking, it is actually the sigmoid function $x \mapsto \frac{1}{1+e^{-x}}$ rather than the shifted sigmoid function $x \mapsto \frac{1}{1+e^{-x}} - \frac{1}{2}$ that is used, e.g., in (Glorot and Bengio, 2010; Mohamed et al., 2011). We incorporated the offset $\frac{1}{2}$ in order to satisfy the requirement $Pf = 0$ for $f = 0$.

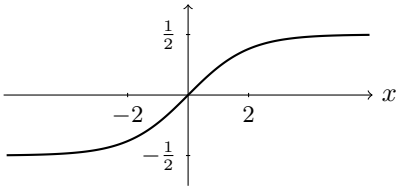

Fig. 2.8: The shifted logistic sigmoid non-linearity on $\mathbb{R}$.

$$= \frac{1}{2}\Big|f(x) - h(x)\Big|. \tag{2.20}$$

It now follows from (2.20) that

$$\begin{aligned}\|Pf - Ph\|_2 &= \Big(\int_{\mathbb{R}^d} |(Pf)(x) - (Ph)(x)|^2 \mathrm{d}x\Big)^{1/2} \\ &\leqslant \frac{1}{2}\Big(\int_{\mathbb{R}^d} |f(x) - h(x)|^2 \mathrm{d}x\Big)^{1/2} = \frac{1}{2}\,\|f - h\|_2,\end{aligned}$$

which establishes Lipschitz continuity of P with $L = \frac{1}{2}$. Since $\mathrm{sig}(0) = 0$, we trivially have $Pf = 0$ for $f = 0$. Finally, (2.14) is satisfied with $\rho(x) := \mathrm{sig}(\mathrm{Re}(x)) + i\,\mathrm{sig}(\mathrm{Im}(x))$.

2.4. POOLING OPERATORS

In the deep learning literature the term "pooling" broadly refers to some form of combining "nearby" values of a signal (e.g., through averaging) or picking one representative value (e.g, through sub-sampling or maximization), see Fig. 2.9. As parts of this thesis (namely, Chapters 3 and 4) deal with DCNNs in *continuous time*[3], it is inevitable to work with continuous-time emulations of discrete-time pooling operators. In this section, we derive these emulations for two *discrete-time*

[3]In the mathematical signal processing literature, the qualifiers *discrete-time* and *continuous-time* allow to differentiate between signals that are i) (square-summable) sequences $f_{\mathrm{d}} \in \ell^2(\mathbb{Z}) := \{f_{\mathrm{d}} : \mathbb{Z} \to \mathbb{C} \mid \sum_{k\in\mathbb{Z}} |f_{\mathrm{d}}[k]|^2 < \infty\}$ and ii) functions $f \in L^2(\mathbb{R}^d)$.

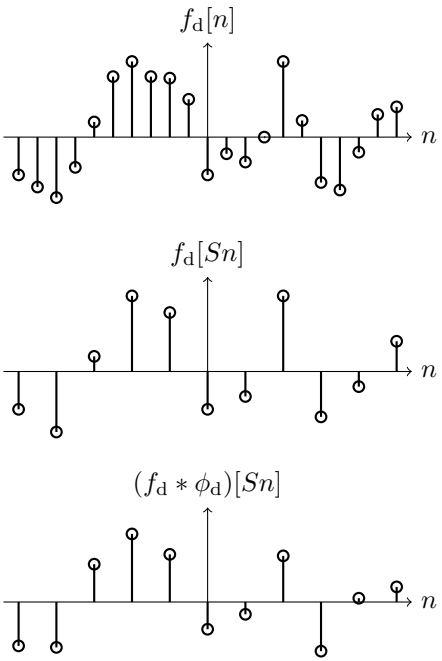

Fig. 2.9: Impact of pooling operators on the discrete-time signal $f_\mathsf{d} \in \ell^2(\mathbb{Z})$ (top row). Pooling by sub-sampling amounts to retaining every S-th sample (middle row). Pooling by averaging (with a box function ϕ_d) amounts to computing local averages of S consecutive samples (bottom row).

pooling operators, namely for pooling by sub-sampling (Pinto et al., 2008) and averaging (Huang and LeCun, 2006; Jarrett et al., 2009).

Pooling by sub-sampling

Consider a one-dimensional discrete-time signal $f_\mathrm{d} \in \ell^2(\mathbb{Z}) := \{f_\mathrm{d} : \mathbb{Z} \to \mathbb{C} \mid \sum_{k\in\mathbb{Z}} |f_\mathrm{d}[k]|^2 < \infty\}$. Sub-sampling by a factor of $S \in \mathbb{N}$ in discrete time is defined by (Vaidyanathan, 1993, Section 4)

$$f_\mathrm{d} \mapsto h_\mathrm{d} := f_\mathrm{d}[S\cdot]$$

and amounts to simply retaining every S-th sample of f_d, see Fig. 2.9 (middle). The discrete-time Fourier transform of h_d is given by a summation over translated and dilated copies of $\widehat{f_\mathrm{d}}$ according to (Vaidyanathan, 1993, Section 4)

$$\widehat{h_\mathrm{d}}(\theta) := \sum_{k\in\mathbb{Z}} h_\mathrm{d}[k] e^{-2\pi i k\theta} = \frac{1}{S}\sum_{k=0}^{S-1} \widehat{f}_\mathrm{d}\Big(\frac{\theta-k}{S}\Big). \tag{2.21}$$

The translated copies of $\widehat{f_\mathrm{d}}$ in (2.21) are a consequence of the 1-periodicity of the discrete-time Fourier transform. We can therefore emulate the discrete-time sub-sampling operator in continuous time through the dilation operator

$$f \mapsto h := S^{d/2} f(S\cdot), \quad f \in L^2(\mathbb{R}^d), \tag{2.22}$$

which in the frequency domain amounts to dilation according to $\widehat{h} = S^{-d/2}\widehat{f}(S^{-1}\cdot)$. The scaling by $S^{d/2}$ in (2.22) ensures unitarity of the continuous-time sub-sampling operator.

Pooling by averaging

In discrete time average pooling is defined by

$$f_\mathrm{d} \mapsto h_\mathrm{d} := (f_\mathrm{d} * \phi_\mathrm{d})[S\cdot] \tag{2.23}$$

for the (typically compactly supported) "averaging kernel" $\phi_\mathrm{d} \in \ell^2(\mathbb{Z})$ and the averaging factor $S \in \mathbb{N}$. Taking ϕ_d to be a box function of length S amounts to computing local averages of S consecutive samples, see Fig. 2.9 (bottom). Weighted averages are obtained by identifying the desired weights with the averaging kernel ϕ_d. The operator (2.23) can be emulated in continuous time according to

$$f \mapsto S^{d/2}\big(f * \phi\big)(S\cdot), \quad f \in L^2(\mathbb{R}^d), \tag{2.24}$$

with the averaging window $\phi \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$.

General pooling

The operators in (2.22) and (2.24) fit into a more general framework. Specifically, we can consider a general pooling operator of the form

$$f \mapsto S^{d/2} P(f)(S\cdot), \tag{2.25}$$

where $S \geqslant 1$ is the so-called pooling factor and $P : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ satisfies the Lipschitz property $\|Pf - Ph\|_2 \leqslant R\|f - h\|_2$, for all $f, h \in L^2(\mathbb{R}^d)$, with $Pf = 0$ for $f = 0$.

The operator in (2.22) can be recovered from (2.25) simply by taking P to equal the identity mapping (which is, of course, Lipschitz-continuous with Lipschitz constant $R = 1$ and satisfies $\mathrm{Id} f = 0$ for $f = 0$). Moreover, (2.24) is recovered from (2.25) by taking $P(f) = f * \phi$, $f \in L^2(\mathbb{R}^d)$, and noting that convolution with ϕ is Lipschitz-continuous with Lipschitz constant $R = \|\phi\|_1$ (thanks to Young's inequality (Grafakos, 2008, Theorem 1.2.12)) and trivially satisfies $Pf = 0$ for $f = 0$.

To make it clear that we consider emulations of discrete-time pooling operators, we refer to the operator in (2.25) as *Lipschitz-pooling through dilation* to indicate that (2.25) essentially amounts to the application of a Lipschitz-continuous mapping followed by a continuous-time dilation.

CHAPTER 3

Deep convolutional feature extraction: Architectures, invariances, and deformation sensitivity

Deep convolutional neural networks have led to breakthrough results in numerous practical machine learning tasks such as classification of images in the ImageNet data set (Krizhevsky et al., 2012; He et al., 2015), control-policy-learning to play Atari games (Mnih et al., 2015) or the board game Go (Silver et al., 2016). Many of these applications first perform feature extraction and then feed the results thereof into a trainable classifier. The mathematical analysis of DCNNs *for feature extraction* was initiated by (Mallat, 2012). Specifically, Mallat considered so-called scattering networks based on a wavelet transform followed by the modulus non-linearity in each network layer, and proved translation invariance (asymptotically in the wavelet scale parameter) and deformation stability of the corresponding feature extractor. This chapter complements Mallat's results by developing a theory of DCNNs for feature extraction encompassing general convolutional transforms, or in more technical parlance, general semi-discrete frames (including Weyl-Heisenberg, curvelet, shearlet, ridgelet, and wavelet frames), general Lipschitz-continuous non-linearities (e.g., rectified linear units, shifted logistic sigmoids, hyperbolic tangents, and modulus functions), and general

Lipschitz-continuous pooling operators emulating sub-sampling and averaging. In addition, all of these elements can be different in different network layers. For the resulting feature extractor we prove a translation invariance result which is of vertical nature in the sense of the network depth determining the amount of invariance, and we establish deformation sensitivity bounds that apply to signal classes with inherent deformation insensitivity such as, e.g., band-limited functions, cartoon functions, and Lipschitz functions.

Outline

The remainder of this chapter is organized as follows. Section 3.1 reviews Mallat's wavelet-based scattering networks. In Section 3.2, we introduce generalized scattering network architectures encompassing general convolutional transforms, general Lipschitz-continuous non-linearities, and general Lipschitz-continuous pooling operators. Section 3.3 contains our first main result, Theorem 1, which shows that the network-based feature extractor is vertical translation-invariant and that pooling plays a crucial role in achieving it. Our second main result, Theorem 2, which provides deformation sensitivity bounds that apply to signal classes with inherent deformation insensitivity (such as, e.g., band-limited functions, cartoon functions, and Lipschitz functions), is presented in Section 3.4. Finally, in Section 3.5, we put our results into perspective and compare them to the results established in (Mallat, 2012).

3.1. MALLAT'S WAVELET-BASED SCATTERING NETWORKS

We set the stage by reviewing scattering networks as introduced in (Mallat, 2012), the basis of which is a multi-layer architecture that involves a wavelet transform followed by the modulus non-linearity, without subsequent pooling. Specifically, (Mallat, 2012, Definition 2.4) defines the feature vector $\Phi_W(f)$ of the signal $f \in L^2(\mathbb{R}^d)$ as the

set[1]

$$\Phi_W(f) := \bigcup_{n=0}^{\infty} \Phi_W^n(f), \tag{3.1}$$

where $\Phi_W^0(f) := \{f * \psi_{(-J,0)}\}$, and

$$\Phi_W^n(f) := \left\{ \left(U\big[\underbrace{\lambda^{(j)}, \ldots, \lambda^{(p)}}_{n \text{ indices}}\big] f\right) * \psi_{(-J,0)} \right\}_{\lambda^{(j)}, \ldots, \lambda^{(p)} \in \Lambda_W \setminus \{(-J,0)\}},$$

for all $n \in \mathbb{N}$, with

$$U\big[\lambda^{(j)}, \ldots, \lambda^{(p)}\big] f := \underbrace{\big| \cdots \big| \, |f * \psi_{\lambda^{(j)}}| * \psi_{\lambda^{(k)}} \big| \cdots * \psi_{\lambda^{(p)}} \big|}_{n-\text{fold convolution followed by modulus}}.$$

Here, the index set $\Lambda_W := \big\{(-J,0)\big\} \cup \big\{(j,k) \mid j \in \mathbb{Z} \text{ with } j > -J,\ k \in \{0, \ldots, K-1\}\big\}$ contains pairs of scales j and directions k (in fact, k is the index of the direction described by the rotation matrix r_k), and

$$\psi_\lambda(x) := 2^{dj} \psi(2^j r_k^{-1} x), \qquad \lambda = (j,k) \in \Lambda_W \setminus \{(-J,0)\}, \tag{3.2}$$

are directional wavelets (Lee, 1996; Antoine et al., 2008; Mallat, 2009) with (complex-valued) mother wavelet $\psi \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$. The r_k, $k \in \{0, \ldots, K-1\}$, are elements of a finite rotation group G (if d is even, G is a subgroup of $SO(d)$; If d is odd, G is a subgroup of $O(d)$). The index $(-J,0) \in \Lambda_W$ is associated with the low-pass filter $\psi_{(-J,0)} \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$, and $J \in \mathbb{Z}$ corresponds to the coarsest scale resolved by the directional wavelets (3.2).

The family of functions $\{\psi_\lambda\}_{\lambda \in \Lambda_W}$ is taken to form a semi-discrete Parseval frame

$$\Psi_{\Lambda_W} := \{T_b I \psi_\lambda\}_{\lambda \in \Lambda_W, b \in \mathbb{R}^d}$$

for $L^2(\mathbb{R}^d)$ (Ali et al., 1993; Kaiser, 1994) and hence satisfies

$$\sum_{\lambda \in \Lambda_W} \int_{\mathbb{R}^d} |\langle f, T_b I \psi_\lambda \rangle|^2 \mathrm{d}b = \sum_{\lambda \in \Lambda_W} \|f * \psi_\lambda\|_2^2 = \|f\|_2^2, \quad \forall f \in L^2(\mathbb{R}^d),$$

[1] We emphasize that the feature vector $\Phi_W(f)$ is a union of the sets of feature vectors $\Phi_W^n(f)$.

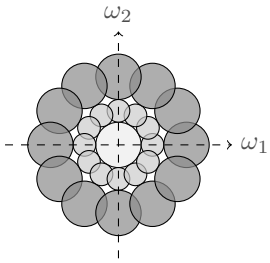

Fig. 3.1: Partitioning of the frequency plane $\mathbb{R}^2$ induced by a semi-discrete directional wavelet frame with $K = 12$ directions.

where $\langle f, T_b I \psi_\lambda \rangle = (f * \psi_\lambda)(b)$, $(\lambda, b) \in \Lambda_{\mathrm{W}} \times \mathbb{R}^d$, are the underlying frame coefficients. Note that for given $\lambda \in \Lambda_{\mathrm{W}}$, we actually have a continuum of frame coefficients as the translation parameter $b \in \mathbb{R}^d$ is left unsampled. We refer to Fig. 3.1 for an illustration of a semi-discrete directional wavelet frame in the frequency domain. In Section 2.2, we give a brief review of the general theory of semi-discrete frames, and in the Sections 2.2.1 and 2.2.2 we collect structured example frames in 1-D and 2-D, respectively.

The architecture corresponding to the feature extractor Φ_W in (3.1), illustrated in Fig. 3.2, is known as *scattering network* (Mallat, 2012), and employs the frame $\Psi_{\Lambda_{\mathrm{W}}}$ and the modulus non-linearity $|\cdot|$ in every network layer, but does not include pooling. For given $n \in \mathbb{N}$, the set $\Phi_W^n(f)$ corresponds to the features of the function f generated in the n-th network layer, see Fig. 3.2.

Remark 3. *The function $|f * \psi_\lambda|$, $\lambda \in \Lambda_W \backslash \{(-J, 0)\}$, can be thought of as indicating the locations of singularities of $f \in L^2(\mathbb{R}^d)$. Specifically, with the relation of $|f * \psi_\lambda|$ to the Canny edge detector (Canny, 1986) as described in (Mallat and Zhong, 1992), in dimension $d = 2$, we can think of $|f * \psi_\lambda| = |f * \psi_{(j,k)}|$, $\lambda = (j, k) \in \Lambda_W \backslash \{(-J, 0)\}$, as an image at scale j specifying the locations of edges of the image f that are oriented in direction k. Furthermore, it was argued in (Bruna and Mallat, 2013; Andén and Mallat, 2014; Oyallon and*

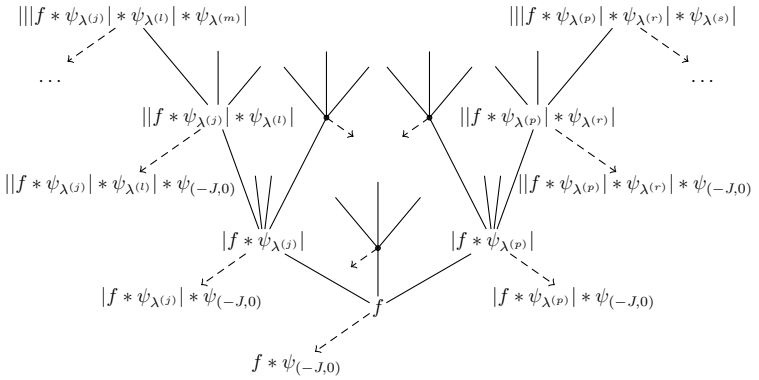

Fig. 3.2: Scattering network architecture based on wavelet filters and the modulus non-linearity. The elements of the feature vector $\Phi_W(f)$ in (3.1) are indicated at the tips of the arrows.

Mallat, 2015) that the feature vector $\Phi_W^1(f)$ generated in the first layer of the scattering network is very similar, in dimension $d = 1$, to mel frequency cepstral coefficients (Davis and Mermelstein, 1980), and in dimension $d = 2$ to SIFT-descriptors (Lowe, 2004; Tola et al., 2010).

It is shown in (Mallat, 2012, Theorem 2.10) that the feature extractor Φ_W is translation-invariant in the sense of

$$\lim_{J\to\infty} |||\Phi_W(T_t f) - \Phi_W(f)||| = 0, \qquad \forall f \in L^2(\mathbb{R}^d), \forall t \in \mathbb{R}^d. \quad (3.3)$$

Note that this invariance result is asymptotic in the scale parameter $J \in \mathbb{Z}$, and does not depend on the network depth, i.e., it guarantees full translation invariance in every network layer. Furthermore, (Mallat, 2012, Theorem 2.12) establishes that Φ_W is stable w.r.t. deformations of the form

$$(F_\tau f)(x) := f(x - \tau(x)),$$

where $\tau : \mathbb{R}^d \to \mathbb{R}^d$. More formally, for the function space $(H_W, \|\cdot\|_{H_W})$ defined in (3.31) below, it is shown in (Mallat, 2012, Theorem 2.12)

that there exists a constant $C > 0$ such that for all $f \in H_W$, and all $\tau \in C^1(\mathbb{R}^d, \mathbb{R}^d)$ with[2] $\|D\tau\|_\infty \leqslant \frac{1}{2d}$, the deformation error satisfies the following deformation stability bound

$$\begin{aligned} &|||\Phi_W(F_\tau f) - \Phi_W(f)||| \\ &\leqslant C\big(2^{-J}\|\tau\|_\infty + J\|D\tau\|_\infty + \|D^2\tau\|_\infty\big)\|f\|_{H_W}. \end{aligned} \tag{3.4}$$

In practice signal classification based on Mallat's wavelet-based scattering networks is performed as follows. First, the function f and the wavelet frame atoms $\{\psi_\lambda\}_{\lambda\in\Lambda_W}$ are discretized to finite-dimensional vectors. The resulting scattering network then computes the finite-dimensional feature vector $\Phi_W(f)$, whose dimension is typically reduced through an orthogonal least squares step (Chen et al., 1991), and then feeds the result into a trainable classifier such as, e.g., a SVM. State-of-the-art results for Mallat's wavelet-based scattering networks were reported for various classification tasks such as handwritten digit recognition (Bruna and Mallat, 2013), texture discrimination (Bruna and Mallat, 2013; Sifre, 2014), and musical genre classification (Andén and Mallat, 2014).

3.2. GENERALIZED SCATTERING NETWORKS

As already mentioned, scattering networks follow the architecture of DCNNs (Rumelhart et al., 1986; LeCun et al., 1990, 1998, 2010, 2015; Serre et al., 2005; Huang and LeCun, 2006; Mutch and Lowe, 2006; Ranzato et al., 2006, 2007; Pinto et al., 2008; Jarrett et al., 2009; Krizhevsky et al., 2012; Bengio et al., 2013) in the sense of cascading convolutions (with atoms $\{\psi_\lambda\}_{\lambda\in\Lambda_W}$ of the wavelet frame Ψ_{Λ_W}) and non-linearities, namely the modulus function, but without pooling. General DCNNs as studied in the literature exhibit a number of additional features:

[2]It is actually the assumption $\|D\tau\|_\infty \leqslant \frac{1}{2d}$, rather than $\|D\tau\|_\infty \leqslant \frac{1}{2}$ as stated in (Mallat, 2012, Theorem 2.12), that is needed in (Mallat, 2012, page 1390) to establish that $|\det(E - (D\tau)(x))| \geqslant 1 - d\|D\tau\|_\infty \geqslant 1/2$.

i) a wide variety of filters are employed, namely pre-specified unstructured filters such as random filters (Ranzato et al., 2007; Jarrett et al., 2009), and filters that are learned in a supervised (Huang and LeCun, 2006; Jarrett et al., 2009) or an unsupervised (Ranzato et al., 2006, 2007; Jarrett et al., 2009) fashion.

ii) a wide variety of non-linearities are used such as, e.g., hyperbolic tangents (Huang and LeCun, 2006; Ranzato et al., 2007; Jarrett et al., 2009), rectified linear units (Nair and Hinton, 2010; Glorot et al., 2011), and logistic sigmoids (Glorot and Bengio, 2010; Mohamed et al., 2011).

iii) convolution and the application of a non-linearity is typically followed by a pooling operator such as, e.g., sub-sampling (Pinto et al., 2008), average-pooling (Huang and LeCun, 2006; Jarrett et al., 2009), or max-pooling (Serre et al., 2005; Mutch and Lowe, 2006; Ranzato et al., 2007; Jarrett et al., 2009).

iv) the filters, non-linearities, and pooling operators are allowed to be different in different network layers (LeCun et al., 2015; Goodfellow et al., 2016).

The purpose of this chapter is to develop a mathematical theory of DCNNs for feature extraction that encompasses all of the aspects above (apart from max-pooling) with the proviso that the pooling operators we analyze are continuous-time emulations of pooling operators in discrete time (see Section 2.4 for the derivation of these emulations). Formally, compared to Mallat's scattering networks, in the n-th network layer, we replace the wavelet-modulus operation $|f * \psi_\lambda|$ by a convolution with the atoms $g_{\lambda_n} \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$ of a general semi-discrete frame $\Psi_n := \{T_b I g_{\lambda_n}\}_{b \in \mathbb{R}^d, \lambda_n \in \Lambda_n}$ for $L^2(\mathbb{R}^d)$ with countable index set Λ_n (see Section 2.2 for a brief review of the theory of semi-discrete frames), followed by a non-linearity $M_n : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ that satisfies the Lipschitz property $\|M_n f - M_n h\|_2 \leqslant L_n \|f - h\|_2$, for all $f, h \in L^2(\mathbb{R}^d)$, with $M_n f = 0$ for $f = 0$. The output of this

non-linearity, $M_n(f * g_{\lambda_n})$, is then pooled according to

$$f \mapsto S_n^{d/2} P_n(f)(S_n \cdot), \tag{3.5}$$

where $S_n \geqslant 1$ is the pooling factor and $P_n : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ satisfies the Lipschitz property $\|P_n f - P_n h\|_2 \leqslant R_n \|f - h\|_2$, for all $f, h \in L^2(\mathbb{R}^d)$, with $P_n f = 0$ for $f = 0$.

We next comment on the different elements in our network architecture in more detail. The frame atoms g_{λ_n} are arbitrary and can, therefore, also be taken to be structured, e.g., Weyl-Heisenberg functions, curvelets, shearlets, ridgelets, or wavelets as considered in (Mallat, 2012) (where the atoms g_{λ_n} are obtained from a mother wavelet through scaling and rotation operations, see Section 3.1). The corresponding semi-discrete signal transforms[3], briefly reviewed in Sections 2.2.1 and 2.2.2, have been employed successfully in various feature extraction tasks (Unser, 1995; Lin and Qu, 2000; Tzanetakis and Cook, 2002; Chen et al., 2005; Arivazhagan et al., 2006; Dettori and Semler, 2007; Ma and Plonka, 2010; Qiao et al., 2010; Ellis et al., 2011), but their use—apart from wavelets—in DCNNs appears to be new. We refer the reader to Section 2.3 for a detailed discussion of several relevant example non-linearities (e.g., rectified linear units, shifted logistic sigmoids, hyperbolic tangents, and, of course, the modulus function) that fit into our framework. Moreover, we refer the reader to Section 2.4 where we explain how the continuous-time pooling operator (3.5) emulates discrete-time pooling operators such as pooling by sub-sampling (Pinto et al., 2008) and averaging (Huang

[3] In the frame literature (Ali et al., 1993; Kaiser, 1994; Candès and Donoho, 2005; Grohs, 2012; Kutyniok and Labate, 2012a; Grohs et al., 2015), a semi-discrete signal transform is a convolutional transform with filters that depend on discrete indices. Specifically, let $\{g_\lambda\}_{\lambda \in \Lambda} \subseteq L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$ be a set of functions indexed by a countable set Λ. Then, the mapping

$$f \mapsto \{f * g_\lambda(b)\}_{\lambda \in \Lambda, b \in \mathbb{R}^d} = \{\langle f, T_b I g_\lambda \rangle\}_{\lambda \in \Lambda, b \in \mathbb{R}^d}, \quad f \in L^2(\mathbb{R}^d), \tag{3.6}$$

is called a semi-discrete signal transform, as it depends on discrete indices $\lambda \in \Lambda$ and continuous variables $b \in \mathbb{R}^d$. We can think of the mapping (3.6) as the analysis operator in frame theory (Daubechies, 1992), with the proviso that for given $\lambda \in \Lambda$, we actually have a continuum of frame coefficients as the translation parameter $b \in \mathbb{R}^d$ is left unsampled.

and LeCun, 2006; Jarrett et al., 2009). As already mentioned in Section 2.4, we refer to the operator in (3.5) as *Lipschitz pooling through dilation* to indicate that (3.5) essentially amounts to the application of a Lipschitz-continuous mapping followed by a continuous-time dilation. We note, however, that the operator in (3.5) will not be unitary in general.

We next state definitions and collect preliminary results needed for the analysis of the general feature extraction network we consider. The basic building blocks of this network are the triplets (Ψ_n, M_n, P_n) associated with individual network layers and referred to as *modules*.

Definition 2. *For $n \in \mathbb{N}$, let $\Psi_n = \{T_b I g_{\lambda_n}\}_{b\in\mathbb{R}^d, \lambda_n\in\Lambda_n}$ be a semi-discrete frame for $L^2(\mathbb{R}^d)$ and let $M_n : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ and $P_n : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ be Lipschitz-continuous operators with $M_n f = 0$ and $P_n f = 0$ for $f = 0$, respectively. Then, the sequence of triplets*

$$\Omega := \big((\Psi_n, M_n, P_n)\big)_{n\in\mathbb{N}}$$

is referred to as a module-sequence.

The following definition introduces the concept of paths on index sets, which will prove helpful in characterizing the feature extraction network. The idea for this formalism is due to (Mallat, 2012).

Definition 3. *Let $\Omega = \big((\Psi_n, M_n, P_n)\big)_{n\in\mathbb{N}}$ be a module-sequence, let $\{g_{\lambda_n}\}_{\lambda_n\in\Lambda_n}$ be the atoms of the frame Ψ_n, and let $S_n \geqslant 1$ be the pooling factor (according to (3.5)) associated with the n-th network layer. Define the operator U_n associated with the n-th layer of the network as $U_n : \Lambda_n \times L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$,*

$$U_n(\lambda_n, f) := U_n[\lambda_n]f := S_n^{d/2} P_n\big(M_n(f * g_{\lambda_n})\big)(S_n\cdot). \tag{3.7}$$

For $1 \leqslant n < \infty$, define the set $\Lambda^n := \Lambda_1 \times \Lambda_2 \times \cdots \times \Lambda_n$. An ordered sequence $q = (\lambda_1, \lambda_2, \ldots, \lambda_n) \in \Lambda^n$ is called a path. For the empty path $e := \emptyset$ we set $\Lambda^0 := \{e\}$ and $U_0[e]f := f$, for all $f \in L^2(\mathbb{R}^d)$.

The operator U_n is well-defined, i.e., $U_n[\lambda_n]f \in L^2(\mathbb{R}^d)$, for all

$(\lambda_n, f) \in \Lambda_n \times L^2(\mathbb{R}^d)$, thanks to

$$\begin{aligned}\|U_n[\lambda_n]f\|_2^2 &= S_n^d \int_{\mathbb{R}^d} \Big|P_n\big(M_n(f * g_{\lambda_n})\big)(S_n x)\Big|^2 \mathrm{d}x \\ &= \int_{\mathbb{R}^d} \Big|P_n\big(M_n(f * g_{\lambda_n})\big)(y)\Big|^2 \mathrm{d}y \\ &= \|P_n\big(M_n(f * g_{\lambda_n})\big)\|_2^2 \leqslant R_n^2 \|M_n(f * g_{\lambda_n})\|_2^2 \qquad (3.8) \\ &\leqslant L_n^2 R_n^2 \|f * g_{\lambda_n}\|_2^2 \leqslant B_n L_n^2 R_n^2 \|f\|_2^2. \qquad (3.9)\end{aligned}$$

For the inequality in (3.8) we used the Lipschitz continuity of P_n according to $\|P_n f - P_n h\|_2^2 \leqslant R_n^2 \|f - h\|_2^2$, together with $P_n h = 0$ for $h = 0$ to get $\|P_n f\|_2^2 \leqslant R_n^2 \|f\|_2^2$. Similar arguments lead to the first inequality in (3.9). The last step in (3.9) is thanks to

$$\|f * g_{\lambda_n}\|_2^2 \leqslant \sum_{\lambda_n' \in \Lambda_n} \|f * g_{\lambda_n'}\|_2^2 \leqslant B_n \|f\|_2^2,$$

which follows from the frame condition (2.1) on Ψ_n. We will also need the extension of the operator U_n to paths $q \in \Lambda^n$ according to

$$U[q]f = U[(\lambda_1, \lambda_2, \ldots, \lambda_n)]f := U_n[\lambda_n] \cdots U_2[\lambda_2] U_1[\lambda_1] f, \qquad (3.10)$$

with $U[e]f := f$. Note that the multi-stage operation (3.10) is again well-defined thanks to

$$\|U[q]f\|_2^2 \leqslant \left(\prod_{k=1}^{n} B_k L_k^2 R_k^2\right) \|f\|_2^2, \qquad \forall q \in \Lambda^n,\ \forall f \in L^2(\mathbb{R}^d), \quad (3.11)$$

which follows by repeated application of (3.9). The signals $U[q]f$, $q \in \Lambda^n$, associated with the n-th network layer, are referred to as feature maps in the deep learning literature.

In scattering networks one atom ψ_λ, $\lambda \in \Lambda_\mathrm{W}$, in the wavelet frame $\Psi_{\Lambda_\mathrm{W}}$, namely the low-pass filter $\psi_{(-J,0)}$, is singled out to generate the extracted features, see Fig. 3.2. We follow this construction and designate one of the atoms in each frame in the module-sequence $\Omega = \big((\Psi_n, M_n, P_n)\big)_{n \in \mathbb{N}}$ as the output-generating atom $\chi_{n-1} := g_{\lambda_n^*}$, $\lambda_n^* \in \Lambda_n$, of the $(n-1)$-th layer. The atoms $\{g_{\lambda_n}\}_{\lambda_n \in \Lambda_n \setminus \{\lambda_n^*\}} \cup \{\chi_{n-1}\}$

in Ψ_n are thus used across two consecutive layers in the sense of $\chi_{n-1} = g_{\lambda_n^*}$ generating the output in the $(n-1)$-th layer, and the $\{g_{\lambda_n}\}_{\lambda_n \in \Lambda_n \backslash \{\lambda_n^*\}}$ propagating signals from the $(n-1)$-th layer to the n-th layer according to (3.7), see Fig. 3.3. Note, however, that the results established in this chapter do not require the output-generating atoms to be low-pass filters[4]. From now on, with slight abuse of notation, we shall write Λ_n for $\Lambda_n \backslash \{\lambda_n^*\}$ as well.

We are now ready to define the feature extractor Φ_Ω based on the module-sequence Ω.

Definition 4. *Let $\Omega = \big((\Psi_n, M_n, P_n)\big)_{n\in\mathbb{N}}$ be a module-sequence. The feature extractor Φ_Ω based on Ω maps $f \in L^2(\mathbb{R}^d)$ to its feature vector*

$$\Phi_\Omega(f) := \bigcup_{n=0}^{\infty} \Phi_\Omega^n(f), \tag{3.12}$$

where $\Phi_\Omega^n(f) := \{(U[q]f) * \chi_n\}_{q\in\Lambda^n}$, for all $n \geqslant 0$.

The set $\Phi_\Omega^n(f)$ in (3.12) corresponds to the features of the function f generated in the n-th network layer, see Fig. 3.3, where $n = 0$ corresponds to the root of the network. The feature extractor $\Phi_\Omega : L^2(\mathbb{R}^d) \to (L^2(\mathbb{R}^d))^{\mathcal{Q}}$, with $\mathcal{Q} := \bigcup_{n=0}^{\infty} \Lambda^n$, is well-defined, i.e., $\Phi_\Omega(f) \in (L^2(\mathbb{R}^d))^{\mathcal{Q}}$, for all $f \in L^2(\mathbb{R}^d)$, under a technical condition on the module-sequence Ω formalized as follows.

Proposition 3. *Let $\Omega = \big((\Psi_n, M_n, P_n)\big)_{n\in\mathbb{N}}$ be a module-sequence. Denote the frame upper bounds of Ψ_n by $B_n > 0$ and the Lipschitz constants of the operators M_n and P_n by $L_n > 0$ and $R_n > 0$, respectively. If*

$$\max\{B_n, B_n L_n^2 R_n^2\} \leqslant 1, \qquad \forall n \in \mathbb{N}, \tag{3.13}$$

then the feature extractor $\Phi_\Omega : L^2(\mathbb{R}^d) \to (L^2(\mathbb{R}^d))^{\mathcal{Q}}$ is well-defined, i.e., $\Phi_\Omega(f) \in (L^2(\mathbb{R}^d))^{\mathcal{Q}}$, for all $f \in L^2(\mathbb{R}^d)$.

Proof. The proof is given in Section 3.6.1. □

[4]It is evident, though, that the actual choices of the output-generating atoms will have an impact on practical performance.

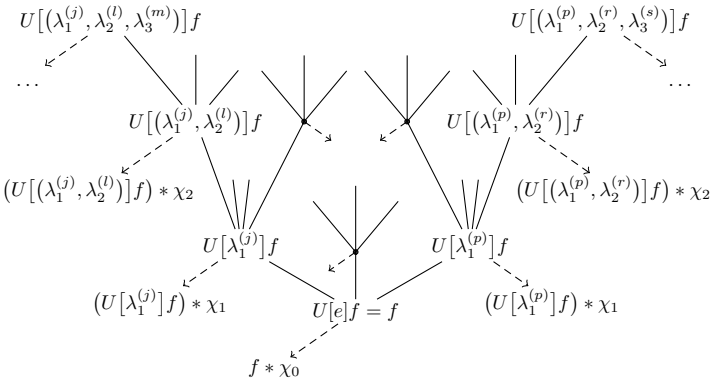

Fig. 3.3: Network architecture underlying the general feature extractor. The index $\lambda_n^{(k)}$ corresponds to the k-th atom $g_{\lambda_n^{(k)}}$ of the frame Ψ_n associated with the n-th network layer. The function χ_n is the output-generating atom of the n-th layer.

As condition (3.13) is of central importance, we formalize it as follows.

Definition 5. *Let $\Omega = \big((\Psi_n, M_n, P_n)\big)_{n\in\mathbb{N}}$ be a module-sequence with frame upper bounds $B_n > 0$ and Lipschitz constants $L_n, R_n > 0$ of the operators M_n and P_n, respectively. The condition*

$$\max\{B_n, B_n L_n^2 R_n^2\} \leqslant 1, \qquad \forall n \in \mathbb{N}, \tag{3.14}$$

is referred to as admissibility condition. Module-sequences that satisfy (3.14) *are called admissible.*

We emphasize that condition (3.14) is easily met in practice. To see this, first note that B_n is determined through the frame Ψ_n (e.g., the directional wavelet frame introduced in Section 3.1 has $B = 1$), L_n is set through the non-linearity M_n (e.g., the modulus function $M = |\cdot|$ has $L = 1$, see Section 2.3), and R_n depends on the operator P_n in (3.5) (e.g., pooling by sub-sampling amounts to $P = \mathrm{Id}$ and has $R = 1$, see Section 2.4). Obviously, condition (3.14) is met if

$$B_n \leqslant \min\{1, L_n^{-2} R_n^{-2}\}, \qquad \forall n \in \mathbb{N},$$

which can be satisfied by simply normalizing the frame elements of Ψ_n accordingly. We refer to Proposition 2 in Section 2.2 for corresponding normalization techniques, which, as explained in the Sections 3.3, 3.4, and 4.3, do not affect our translation invariance result, our deformation sensitivity bounds, as well as our energy decay and conservation results.

3.3. VERTICAL TRANSLATION INVARIANCE

The following theorem states that under very mild decay conditions on the Fourier transforms $\widehat{\chi_n}$ of the output-generating atoms χ_n, the feature extractor Φ_Ω exhibits vertical translation invariance in the sense of the features becoming more translation-invariant with increasing network depth. This result is in line with observations made in the deep learning literature, e.g., in (Serre et al., 2005; Huang and LeCun, 2006; Mutch and Lowe, 2006; Ranzato et al., 2007; Jarrett et al., 2009), where it is informally argued that the network outputs generated at deeper layers tend to be more translation-invariant.

Theorem 1. *Let $\Omega = \big((\Psi_n, M_n, P_n)\big)_{n\in\mathbb{N}}$ be an admissible module-sequence, let $S_n \geqslant 1$, $n \in \mathbb{N}$, be the pooling factors in* (3.7)*, and assume that the operators $M_n : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ and $P_n : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ commute with the translation operator T_t, i.e.,*

$$M_n T_t f = T_t M_n f, \qquad P_n T_t f = T_t P_n f, \tag{3.15}$$

for all $f \in L^2(\mathbb{R}^d)$, all $t \in \mathbb{R}^d$, and all $n \in \mathbb{N}$.

i) The features $\Phi_\Omega^n(f)$ generated in the n-th network layer satisfy

$$\Phi_\Omega^n(T_t f) = T_{t/(S_1 \cdots S_n)} \Phi_\Omega^n(f), \tag{3.16}$$

for all $f \in L^2(\mathbb{R}^d)$, all $t \in \mathbb{R}^d$, and all $n \in \mathbb{N}$. Here, $T_t \Phi_\Omega^n(f)$ refers to element-wise application of T_t, i.e.,

$$T_t \Phi_\Omega^n(f) := \{T_t h \,|\, h \in \Phi_\Omega^n(f)\}.$$

ii) If, in addition, there exists a constant $K > 0$ (that does not depend on n) such that the Fourier transforms $\widehat{\chi_n}$ of the output-generating atoms χ_n satisfy the decay condition

$$|\widehat{\chi_n}(\omega)||\omega| \leqslant K, \qquad \text{a.e. } \omega \in \mathbb{R}^d, \ \forall n \in \mathbb{N}_0, \tag{3.17}$$

then

$$|||\Phi_\Omega^n(T_t f) - \Phi_\Omega^n(f)||| \leqslant \frac{2\pi|t|K}{S_1 \cdots S_n}\|f\|_2, \tag{3.18}$$

for all $f \in L^2(\mathbb{R}^d)$, all $t \in \mathbb{R}^d$, and all $n \in \mathbb{N}$.

Proof. The proof is given in Section 3.6.2. □

We start by noting that all pointwise non-linearities $M_n : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ satisfy the commutation relation in (3.15). A large class of non-linearities widely used in the deep learning literature, such as rectified linear units, hyperbolic tangents, shifted logistic sigmoids, and the modulus function as employed in (Mallat, 2012), are, indeed, pointwise and hence covered by Theorem 1. Moreover, $P = \mathrm{Id}$ as in pooling by sub-sampling trivially satisfies (3.15). Pooling by averaging $Pf = f * \phi$, with $\phi \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$, satisfies (3.15) as a consequence of the convolution operator commuting with the translation operator T_t. Note that (3.17) can easily be met by taking the output-generating atoms $\{\chi_n\}_{n\in\mathbb{N}_0}$ either to satisfy

$$\sup_{n\in\mathbb{N}_0}\{\|\chi_n\|_1 + \|\nabla\chi_n\|_1\} < \infty, \tag{3.19}$$

see, e.g., (Rudin, 1991, Chapter 7), or to be uniformly band-limited in the sense of $\mathrm{supp}(\widehat{\chi_n}) \subseteq B_r(0)$, for all $n \in \mathbb{N}_0$, with an r that is independent of n (see, e.g., (Mallat, 2009, Chapter 2.3)).

The bound in (3.18) shows that we can explicitly control the amount of translation invariance via the pooling factors S_n. This result is in line with observations made in the deep learning literature, e.g., in (Serre et al., 2005; Huang and LeCun, 2006; Mutch and Lowe, 2006; Ranzato et al., 2007; Jarrett et al., 2009), where it is informally argued that pooling is crucial to get translation invariance of the extracted

(a) (b) (c)

Fig. 3.4: Handwritten digits from the MNIST data set (LeCun and Cortes, 1998). For practical machine learning tasks (e.g., signal classification), we often want the feature vector $\Phi_\Omega(f)$ to be invariant to the digits' spatial location within the image f. Theorem 1 establishes that the features $\Phi_\Omega^n(f)$ become more translation-invariant with increasing layer index n.

features. Furthermore, the condition $\lim_{n\to\infty} S_1 \cdot S_2 \cdot \ldots \cdot S_n = \infty$ (easily met by taking $S_n > 1$, for all $n \in \mathbb{N}$) guarantees, thanks to (3.18), asymptotically full translation invariance according to

$$\lim_{n\to\infty} |||\Phi_\Omega^n(T_t f) - \Phi_\Omega^n(f)||| = 0, \qquad \forall f \in L^2(\mathbb{R}^d), \forall t \in \mathbb{R}^d. \quad (3.20)$$

This means that the features $\Phi_\Omega^n(T_t f)$ corresponding to the shifted versions $T_t f$ of the handwritten digit "3" in Figs. 3.4 (b) and (c) with increasing network depth increasingly "look like" the features $\Phi_\Omega^n(f)$ corresponding to the unshifted handwritten digit in Fig. 3.4 (a). Casually speaking, the shift operator T_t is increasingly absorbed by Φ_Ω^n as $n \to \infty$, with the upper bound (3.18) quantifying this absorption w.r.t. the layer index n, the constant K, and the pooling factors $\{S_k\}_{k=1}^n$. In contrast, the translation invariance result (3.3) established in (Mallat, 2012) is asymptotic in the wavelet scale parameter J, and does not depend on the network depth, i.e., it guarantees full translation invariance in every network layer. We honor this difference by referring to (3.3) as *horizontal* translation invariance and to (3.20) as *vertical* translation invariance.

We emphasize that vertical translation invariance is a structural property. Specifically, if P_n is unitary (such as, e.g., in the case of pooling by sub-sampling where P_n simply equals the identity

mapping), then so is the pooling operator in (3.5) owing to

$$\begin{aligned}\|S_n^{d/2}P_n(f)(S_n\cdot)\|_2^2 &= S_n^d\int_{\mathbb{R}^d}|P_n(f)(S_nx)|^2\mathrm{d}x = \int_{\mathbb{R}^d}|P_n(f)(x)|^2\mathrm{d}x\\ &= \|P_n(f)\|_2^2 = \|f\|_2^2,\end{aligned}$$

where we employed the change of variables $y = S_nx$, $\frac{\mathrm{d}y}{\mathrm{d}x} = S_n^d$.

Finally, we note that in practice in certain applications it is actually translation *covariance* in the sense of $\Phi_\Omega^n(T_tf) = T_t\Phi_\Omega^n(f)$, for all $f \in L^2(\mathbb{R}^d)$ and all $t \in \mathbb{R}^d$, that is desirable, for example, in facial landmark detection where the goal is to estimate the absolute position of facial landmarks in images. In such applications features in the layers closer to the root of the network are more relevant as they are less translation-invariant and more translation-covariant. The reader is referred to Section 5.6 where corresponding numerical evidence is provided. We proceed to the formal statement of our translation covariance result.

Corollary 1. *Let $\Omega = \big((\Psi_n, M_n, P_n)\big)_{n\in\mathbb{N}}$ be an admissible module-sequence, let $S_n \geqslant 1$, $n \in \mathbb{N}$, be the pooling factors in* (3.7)*, and assume that the operators $M_n : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ and $P_n : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ commute with the translation operator T_t in the sense of* (3.15)*. If, in addition, there exists a constant $K > 0$ (that does not depend on n) such that the Fourier transforms $\widehat{\chi_n}$ of the output-generating atoms χ_n satisfy the decay condition* (3.17)*, then*

$$|||\Phi_\Omega^n(T_tf) - T_t\Phi_\Omega^n(f)||| \leqslant 2\pi|t|K\big|1/(S_1\dots S_n) - 1\big|\|f\|_2,$$

for all $f \in L^2(\mathbb{R}^d)$, all $t \in \mathbb{R}^d$, and all $n \in \mathbb{N}$.

Proof. The proof is given in Section 3.6.3. □

Corollary 1 shows that no pooling, i.e., taking $S_n = 1$, for all $n \in \mathbb{N}$, leads to full translation covariance in every network layer. Conversely, this proves that pooling is necessary to get vertical translation invariance as otherwise the features remain fully translation-covariant irrespective of the network depth.

(a) (b) (c)

Fig. 3.5: Handwritten digits from the MNIST data set (LeCun and Cortes, 1998). If f denotes the image of the handwritten digit "5" in (a), then—for appropriately chosen τ—the function $F_\tau f = f(\cdot - \tau(\cdot))$ models images of "5" based on different handwriting styles as in (b) and (c).

3.4. DEFORMATION SENSITIVITY BOUNDS

In this section we provide bounds on the sensitivity of the feature extractor Φ_Ω w.r.t. deformations of the form

$$(F_\tau f)(x) := f(x - \tau(x)).$$

This class of deformations encompasses non-linear distortions $f(x - \tau(x))$ as illustrated in Fig. 3.5, inter alia.

3.4.1. Decoupling

The deformation sensitivity bounds we derive are signal-class specific in the sense of applying to input signals taken from a particular class. Specifically, the signal class needs to exhibit inherent deformation insensitivity in the following sense.

Definition 6. *A signal class $\mathcal{C} \subseteq L^2(\mathbb{R}^d)$ is called deformation-insensitive if there exist $\alpha, C > 0$ such that for all $f \in \mathcal{C}$ and all (possibly non-linear) $\tau \in C^1(\mathbb{R}^d, \mathbb{R}^d)$ with $\|\tau\|_\infty < \frac{1}{2}$ and $\|D\tau\|_\infty \leqslant \frac{1}{2d}$, it holds that*

$$\|f - F_\tau f\|_2 \leqslant C\|\tau\|_\infty^\alpha \|f\|_2. \tag{3.21}$$

The constant $C > 0$ and the Lipschitz exponent $\alpha > 0$ in (3.21) depend on the particular signal class $\mathcal{C}$. Moreover, $\alpha > 0$ determines

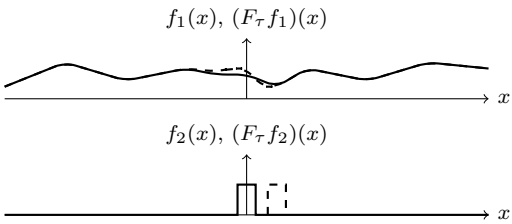

Fig. 3.6: Impact of the deformation F_τ, with $\tau(x) = \frac{1}{2}\, e^{-x^2}$, on the functions $f_1 \in \mathcal{C}_1 \subseteq L^2(\mathbb{R})$ and $f_2 \in \mathcal{C}_2 \subseteq L^2(\mathbb{R})$. The signal class $\mathcal{C}_1$ consists of smooth, slowly varying functions (e.g., band-limited functions, see Section 3.4.2), and $\mathcal{C}_2$ consists of compactly supported functions that exhibit discontinuities (e.g., cartoon functions, see Section 3.4.3). We observe that f_1, unlike f_2, is affected only mildly by F_τ. The amount of deformation induced therefore depends drastically on the specific $f \in L^2(\mathbb{R})$.

the decay rate of the deformation error $\|f - F_\tau f\|_2$ as $\|\tau\|_\infty \to 0$. Clearly, larger $\alpha > 0$ results in the deformation error decaying faster as the deformation becomes smaller. Examples of deformation-insensitive signal classes are

i) the class of band-limited functions with $\alpha = 1$ (see Section 3.4.2),

ii) the class of cartoon functions with $\alpha = \frac{1}{2}$ (see Section 3.4.3),

iii) the class of Lipschitz functions with $\alpha = 1$ (see Section 3.4.4).

While a deformation sensitivity bound that applies to all $f \in L^2(\mathbb{R}^d)$ would be desirable, the example in Fig. 3.6 illustrates the difficulty underlying this desideratum. Specifically, we can see in Fig. 3.6 that for given $\tau(x)$ the impact of the deformation induced by $f(x - \tau(x))$ can depend drastically on the function $f \in L^2(\mathbb{R}^d)$ itself. We note that the deformation stability bound (3.4) for scattering networks reported in (Mallat, 2012, Theorem 2.12) applies to a signal class as well, see (3.31) in Section 3.5.

Remark 4. *It is interesting to note that in order to obtain bounds of the form* $\|f - F_\tau f\|_2 \leq C\|\tau\|_\infty^\alpha \|f\|_2$*, for* $f \in \mathcal{C} \subseteq L^2(\mathbb{R}^d)$*, for some*

$C > 0$ and some $\alpha > 0$, we need to impose non-trivial constraints on the set $\mathcal{C} \subseteq L^2(\mathbb{R}^d)$. Indeed, consider $d = 1$ and $\tau_s(x) = s$, for some $s < \frac{1}{2}$; the corresponding deformation F_{τ_s} amounts to a simple translation by s with $\|\tau_s\|_\infty = s < \frac{1}{2}$ and $\|D\tau_s\|_\infty = 0 \leqslant \frac{1}{2d}$. Let $f_s \in L^2(\mathbb{R}^d)$ be a function that has its energy $\|f_s\|_2^2 = 1$ concentrated in a small interval according to $supp(f_s) \subseteq [-s/2, s/2]$. Then, f_s and $F_{\tau_s} f_s$ have disjoint support sets and hence $\|f_s - F_{\tau_s} f_s\|_2 = \sqrt{2}$, which does not decay with $\|\tau\|_\infty^\alpha = s^\alpha$ for any $\alpha > 0$.

Our signal-class specific deformation sensitivity bound for the feature extractor Φ_Ω is based on the following two ingredients. First, we establish—in Proposition 7 in Section 3.6.8—that the feature extractor Φ_Ω is Lipschitz-continuous with Lipschitz constant $L_\Omega = 1$, i.e.,

$$|||\Phi_\Omega(f) - \Phi_\Omega(h)||| \leqslant \|f - h\|_2, \qquad \forall f, h \in L^2(\mathbb{R}^d). \tag{3.22}$$

Second, we derive for the signal classes under consideration (namely, for band-limited functions in Section 3.4.2, for cartoon functions in Section 3.4.3, and for Lipschitz functions in Section 3.4.4) an upper bound on the deformation error $\|f - F_\tau f\|_2$ according to (3.21). The deformation sensitivity bound for the feature extractor is then obtained by setting $h = F_\tau f$ in (3.22) and using (3.21) (see Section 3.6.4 for the corresponding technical details). This "decoupling" into Lipschitz continuity of Φ_Ω and a deformation sensitivity bound for the underlying signal class has important practical ramifications as it shows that whenever we have a deformation sensitivity bound for a signal class, we automatically get a deformation sensitivity bound for the corresponding feature extractor thanks to its Lipschitz continuity. We proceed to the formal statement of the deformation sensitivity result.

Theorem 2. *Let $\Omega = \big((\Psi_n, M_n, P_n)\big)_{n\in\mathbb{N}}$ be an admissible module-sequence and let $\mathcal{C} \subseteq L^2(\mathbb{R}^d)$ be a deformation-insensitive signal class. There exist constants $\alpha, C > 0$ (that do not depend on Ω) such that for all $f \in \mathcal{C}$ and all $\tau \in C^1(\mathbb{R}^d, \mathbb{R}^d)$ with $\|\tau\|_\infty < \frac{1}{2}$ and $\|D\tau\|_\infty \leqslant \frac{1}{2d}$,*

the feature extractor Φ_Ω *satisfies*

$$|||\Phi_\Omega(F_\tau f) - \Phi_\Omega(f)||| \leqslant C\|\tau\|_\infty^\alpha \|f\|_2. \tag{3.23}$$

Proof. The proof is given in Section 3.6.4. □

First, we note that the bound in (3.23) holds for sufficiently "small" τ, i.e., as long as $\|\tau\|_\infty < \frac{1}{2}$ and $\|D\tau\|_\infty \leqslant \frac{1}{2d}$. We can think of this condition on τ and on the Jacobian matrix $D\tau$ as follows: Let f be an image of the handwritten digit "5" (see Fig. 3.5 (a)). Then, $\{F_\tau f \mid \|\tau\|_\infty < \frac{1}{2} \text{ and } \|D\tau\|_\infty \leqslant \frac{1}{2d}\}$ is a collection of images of the handwritten digit "5", where each $F_\tau f$ models an image that may be generated, e.g., based on a different handwriting style (see Figs. 3.5 (b) and (c)). The bounds $\|\tau\|_\infty < \frac{1}{2}$ and $\|D\tau\|_\infty \leqslant \frac{1}{2d}$ now impose a quantitative limit on the amount of deformation tolerated. The deformation sensitivity bound (3.23) provides a limit on how much the features corresponding to the images in the set $\{F_\tau f \mid \|\tau\|_\infty < \frac{1}{2} \text{ and } \|D\tau\|_\infty \leqslant \frac{1}{2d}\}$ can differ. The strength of the deformation sensitivity bound in Theorem 2 derives itself from the fact that the only condition on the underlying module-sequence Ω needed is admissibility according to (3.14), which as outlined in Section 3.2, can easily be obtained by normalizing the frame elements of Ψ_n, for all $n \in \mathbb{N}$, appropriately. This normalization does not have an impact on the constant C in (3.23). More specifically, C is shown in Section 3.6.4 to be completely independent of Ω. All this is thanks to the decoupling technique used to prove Theorem 2 being completely independent of the structures of the frames Ψ_n and of the specific form of the Lipschitz-continuous operators M_n and P_n. Moreover, as the vertical translation invariance result in Theorem 1 in Section 3.3 applies to all $f \in L^2(\mathbb{R}^d)$, the results established in this chapter show that vertical translation invariance and limited sensitivity to deformations—for signal classes with inherent deformation insensitivity—are guaranteed by the network structure per se rather than the specific convolution filters, non-linearities, and pooling operators.

Finally, we note that the bound (3.4) for scattering networks reported in (Mallat, 2012, Theorem 2.12) depends upon first-order $(D\tau)$

and second-order ($D^2\tau$) derivatives of τ. In contrast, our bound (3.23) depends on ($D\tau$) implicitly only as we need to impose the condition $\|D\tau\|_\infty \leqslant \frac{1}{2d}$ for the bound to hold[5]. We honor this difference by referring to (3.4) as deformation *stability* bound and to our bound (3.23) as deformation *sensitivity* bound.

Remark 5. *It is interesting to note that the frame lower bounds $A_n > 0$ of the semi-discrete frames Ψ_n affect neither the vertical translation invariance result in Theorem 1 in Section 3.3 nor the deformation sensitivity bound in Theorem 2. In fact, our entire theory carries through as long as the collections $\Psi_n = \{T_b I g_{\lambda_n}\}_{\lambda_n \in \Lambda_n, b \in \mathbb{R}^d}$, for all $n \in \mathbb{N}$, satisfy the Bessel property*

$$\sum_{\lambda_n \in \Lambda_n} \int_{\mathbb{R}^d} |\langle f, T_b I g_{\lambda_n} \rangle|^2 \mathrm{d}b = \sum_{\lambda_n \in \Lambda_n} \|f * g_{\lambda_n}\|_2^2 \leqslant B_n \|f\|_2^2,$$

for all $f \in L^2(\mathbb{R}^d)$, for some $B_n > 0$, which, by Proposition 1 in Section 2.2, is equivalent to

$$\sum_{\lambda_n \in \Lambda_n} |\widehat{g_{\lambda_n}}(\omega)|^2 \leqslant B_n, \qquad \text{a.e. } \omega \in \mathbb{R}^d. \tag{3.24}$$

Pre-specified unstructured filters (Ranzato et al., 2007; Jarrett et al., 2009) and learned filters (Huang and LeCun, 2006; Ranzato et al., 2006, 2007; Jarrett et al., 2009) are therefore covered by our theory as long as (3.24) *is satisfied. In classical frame theory $A_n > 0$ guarantees completeness of the set $\Psi_n = \{T_b I g_{\lambda_n}\}_{\lambda_n \in \Lambda_n, b \in \mathbb{R}^d}$ for the signal space under consideration, here $L^2(\mathbb{R}^d)$. The absence of a frame lower bound $A_n > 0$ therefore translates into a lack of completeness of Ψ_n, which may result in the frame coefficients $\langle f, T_b I g_{\lambda_n} \rangle = (f * g_{\lambda_n})(b)$, $(\lambda_n, b) \in \Lambda_n \times \mathbb{R}^d$, not containing all essential features of the signal f. This will, in general, have a (possibly significant) impact on practical feature extraction performance which is why ensuring the entire frame property* (2.1) *is prudent. Interestingly, satisfying the frame property* (2.1) *for all Ψ_n, $n \in \mathbb{Z}$, does, however, not guarantee that the feature*

[5] We note that $\|D\tau\|_\infty \leqslant \frac{1}{2d}$ is needed for the bound (3.4) to hold as well.

extractor Φ_Ω has a trivial null-set, i.e., $\Phi_\Omega(f) = 0$ if and only if $f = 0$. We refer the reader to Section 4.6 for an example of a feature extractor with non-trivial null-set.

3.4.2. Bounds for band-limited functions

The following proposition states that the signal class of L-band-limited functions

$$L_L^2(\mathbb{R}^d) := \big\{ f \in L^2(\mathbb{R}^d) \mid \operatorname{supp}(\widehat{f}\,) \subseteq B_L(0) \big\}, \quad L > 0,$$

exhibits inherent deformation insensitivity in the sense of Definition 6 in Section 3.4.1.

Proposition 4. *There exists a constant $C > 0$ such that for all $f \in L_R^2(\mathbb{R}^d)$ and all $\tau \in C^1(\mathbb{R}^d, \mathbb{R}^d)$ with $\|D\tau\|_\infty \leqslant \frac{1}{2d}$, it holds that*

$$\|f - F_\tau f\|_2 \leqslant CL\|\tau\|_\infty \|f\|_2. \tag{3.25}$$

Proof. The proof is given in Section 3.6.5. □

The dependence of the upper bound in (3.25) on the bandwidth L reflects the intuition that the deformation sensitivity bound should depend on the input signal class "description complexity". Many signals of practical significance (e.g., natural images, see Fig. 3.7) are, however, either not band-limited due to the presence of sharp (and possibly curved) edges or exhibit large bandwidths. In the latter case, the bound (3.25) effectively becomes void owing to its linear dependence on L. We refer the reader to Section 3.4.3 where deformation sensitivity bounds for non-smooth signals are established.

A similar bound to (3.25) was derived in (Mallat, 2012, Appendix B) for wavelet-based scattering networks, namely

$$\|f * \psi_{(-J,0)} - F_\tau(f * \psi_{(-J,0)})\|_2 \leqslant C 2^{-J+d} \|\tau\|_\infty \|f\|_2, \tag{3.26}$$

for all $f \in L^2(\mathbb{R}^d)$, where $\psi_{(-J,0)}$ is the low-pass filter of a semi-discrete directional wavelet frame for $L^2(\mathbb{R}^d)$. The techniques for proving (3.25)

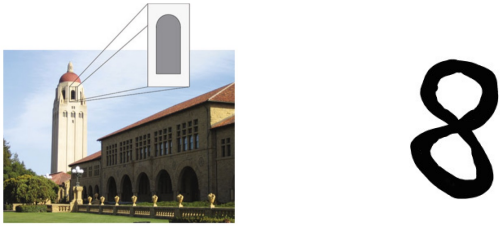

Fig. 3.7: Left: A natural image (image credit: (Kutyniok and Labate, 2012a)) is typically governed by areas of little variation, with the individual areas separated by edges that can be modeled as curved singularities. Right: An image of a handwritten digit.

and (3.26) are related in the sense of both employing Schur's Lemma (Grafakos, 2008, Appendix I.1) and a Taylor series expansion argument (Rudin, 1983, page 411). The signal-class specificity of our bound (3.25) comes with new technical elements detailed at the beginning of the proof in Section 3.6.5.

3.4.3. Bounds for cartoon functions

As already mentioned, the bound in (3.25) applies to the space of L-band-limited functions. Many signals of practical significance (e.g., natural images) are, however, not band-limited (due to the presence of sharp and possibly curved edges, see Fig. 3.7) or exhibit large bandwidths. In the latter case, the deformation sensitivity bound (3.25) becomes void as it depends linearly on L. The goal of this section is to take structural properties of natural images into account by considering the class of cartoon functions introduced in (Donoho, 2001). These functions satisfy mild decay properties and are piecewise continuously differentiable apart from curved discontinuities along C^2-hypersurfaces. Cartoon functions provide a good model for natural images (see Fig. 3.7, left) such as those in the Caltech-256 (Griffin et al., 2007) and CIFAR-100 (Krizhevsky, 2009) data sets, for images of handwritten digits (LeCun and Cortes, 1998) (see Fig. 3.7, right),

and for images of geometric objects of different shapes, sizes, and colors as in the Baby AI School data set[6].

We will work with the following—relative to the definition in (Donoho, 2001)—slightly modified version of cartoon functions.

Definition 7. *The function $f : \mathbb{R}^d \to \mathbb{C}$ is referred to as a cartoon function if it can be written as $f = f_1 + \mathbb{1}_B f_2$, where $B \subseteq \mathbb{R}^d$ is a compact domain whose boundary ∂B is a compact topologically embedded C^2-hypersurface of $\mathbb{R}^d$ without boundary[7], and $f_i \in H^{1/2}(\mathbb{R}^d) \cap C^1(\mathbb{R}^d, \mathbb{C})$, $i = 1, 2$, satisfy the decay condition*

$$|\nabla f_i(x)| \leq C\langle x\rangle^{-d}, \quad i = 1, 2, \tag{3.27}$$

for some $C > 0$ (not depending on f_1,f_2). Furthermore, we denote by

$$\begin{aligned}\mathcal{C}^K_{\mathrm{CART}} := \{&f_1 + \mathbb{1}_B f_2 \mid f_i \in H^{1/2}(\mathbb{R}^d) \cap C^1(\mathbb{R}^d, \mathbb{C}),\\ &|\nabla f_i(x)| \leq K\langle x\rangle^{-d},\ \mathrm{vol}^{d-1}(\partial B) \leqslant K,\ \|f_2\|_\infty \leqslant K\}\end{aligned}$$

the class of cartoon functions of "size" $K > 0$.

We chose the term "size" to indicate the length $\mathrm{vol}^{d-1}(\partial B)$ of the hypersurface ∂B. Furthermore, $\mathcal{C}^K_{\mathrm{CART}} \subseteq L^2(\mathbb{R}^d)$, for all $K > 0$. This simply follows from the triangle inequality according to $\|f_1 + \mathbb{1}_B f_2\|_2 \leqslant \|f_1\|_2 + \|\mathbb{1}_B f_2\|_2 \leqslant \|f_1\|_2 + \|f_2\|_2 < \infty$, where in the last step we used $f_1, f_2 \in H^{1/2}(\mathbb{R}^d) \subseteq L^2(\mathbb{R}^d)$. Finally, we note that our results can easily be generalized to finite linear combinations of cartoon functions, but this is not done here for simplicity of exposition.

We proceed to the formal statement of our deformation insensitivity result.

Proposition 5. *For every $K > 0$ there exists a constant $C_K > 0$ such that for all $f \in \mathcal{C}^K_{\mathrm{CART}}$ and all (possibly non-linear) $\tau : \mathbb{R}^d \to \mathbb{R}^d$ with $\|\tau\|_\infty < \frac{1}{2}$, it holds that*

$$\|f - F_\tau f\|_2 \leq C_K \|\tau\|_\infty^{1/2}. \tag{3.28}$$

[6] `http://www.iro.umontreal.ca/%7Elisa/twiki/bin/view.cgi/Public/BabyAISchool`

[7] We refer the reader to (do Carmo, 2013, Chapter 0) for a review on differentiable manifolds.

Proof. The proof is given in Section 3.6.6. □

The dependence of C_K on K reflects the intuition that the deformation sensitivity bound should depend on the signal class description complexity. For band-limited signals, this dependence is exhibited by the right hand side (RHS) in (3.25) being linear in the bandwidth L. The Lipschitz exponent $\alpha = \frac{1}{2}$ on the RHS of (3.28) determines the decay rate of the deformation error $\|f - F_\tau f\|_2$ as $\|\tau\|_\infty \to 0$. Clearly, larger $\alpha > 0$ results in the deformation error decaying faster as the deformation becomes smaller. The following simple example shows that the Lipschitz exponent $\alpha = \frac{1}{2}$ in (3.28) is best possible, i.e., it can not be larger. Consider, again, $d = 1$ and $\tau_s(x) = s$, for a fixed s satisfying $0 < s < \frac{1}{2}$. Let $f = \mathbb{1}_{[-1,1]}$. Then $f \in \mathcal{C}^K_{\mathrm{CART}}$ for some $K > 0$ and $\|f - F_{\tau_s} f\|_2 = \sqrt{2s} = \sqrt{2}\|\tau\|_\infty^{1/2}$.

3.4.4. Bounds for Lipschitz functions

The following proposition states that functions f that do not exhibit discontinuities along C^2-hypersurfaces (such as cartoon functions), but otherwise satisfy the decay condition (3.27), are deformation-insensitive. More formally, we establish (3.21) with $\alpha = 1$ for the signal class

$$V_R := \left\{ f \in L^2(\mathbb{R}^d) \cap C^1(\mathbb{R}^d, \mathbb{C}) \mid |\nabla f(x)| \leq R\langle x \rangle^{-d} \right\}, \quad R > 0.$$

Proposition 6. *For every $R > 0$ there exists a constant $C_R > 0$ such that for all $f \in V_R$ and all (possibly non-linear) $\tau : \mathbb{R}^d \to \mathbb{R}^d$ with $\|\tau\|_\infty < \frac{1}{2}$, it holds that*

$$\|f - F_\tau f\|_2 \leq C_R \|\tau\|_\infty. \tag{3.29}$$

Proof. The proof is given in Section 3.6.7. □

We note that the condition $f \in C^1(\mathbb{R}^d, \mathbb{C})$ in Proposition 6 can be relaxed to Lipschitz-continuous f. This follows simply by noting that Lipschitz-continuous functions are differentiable a.e. (Federer, 1969, Theorem 3.1.6) and that, since we only bound L^2-norms, sets of Lebesgue measure zero can be ignored.

3.5. RELATION TO MALLAT'S RESULTS

3.5.1. Architectures

To see how Mallat's wavelet-modulus feature extractor Φ_W defined in (3.1) is covered by our generalized framework, simply note that Φ_W is a feature extractor Φ_Ω based on the module-sequence

$$\Omega_W = \big((\Psi_{\Lambda_W}, |\cdot|, \mathrm{Id})\big)_{n\in\mathbb{N}},$$

where each layer is associated with the same module $(\Psi_{\Lambda_W}, |\cdot|, \mathrm{Id})$ and thus with the same semi-discrete directional wavelet frame $\Psi_{\Lambda_W} = \{T_b I \psi_\lambda\}_{\lambda\in\Lambda_W, b\in\mathbb{R}^d}$ and the modulus non-linearity $|\cdot|$. Since Φ_W does not involve pooling, we have $P_n = \mathrm{Id}$ and $S_n = 1$, for all $n \in \mathbb{N}$. The output-generating atom for all layers is taken to be the low-pass filter $\psi_{(-J,0)}$, i.e., $\chi_n = \psi_{(-J,0)}$, for all $n \in \mathbb{N}_0$. Owing to (Mallat, 2012, Equation 2.7), the set $\{\psi_\lambda\}_{\lambda\in\Lambda_W}$ satisfies the equivalent frame condition (2.3) with $A = B = 1$, and Ψ_{Λ_W} therefore forms a semi-discrete Parseval frame for $L^2(\mathbb{R}^d)$, which implies $A_n = B_n = 1$, for all $n \in \mathbb{N}$. The modulus non-linearity $M_n = |\cdot|$ and the operator $P_n = \mathrm{Id}$ are Lipschitz-continuous with Lipschitz constants $L_n = 1$ and $R_n = 1$, and satisfy $M_n f = |f| = 0$ and $P_n f = f = 0$ for $f = 0$, respectively. Therefore, the weak admissibility condition (3.14) is met according to

$$\max\{B_n, B_n R_n^{-d} L_n^2\} = \max\{1, 1\} = 1 \leqslant 1, \qquad \forall n \in \mathbb{N}. \tag{3.30}$$

Moreover, $M_n = |\cdot|$ and $P_n = \mathrm{Id}$ trivially commute with the translation operator T_t in the sense of (3.15), see Section 2.3 for the corresponding formal arguments. Owing to $|\psi_{(-J,0)}(x)| \leqslant C_1(1+|x|)^{-d-2}$ and $|\nabla\psi_{(-J,0)}(x)| \leqslant C_2(1+|x|)^{-d-2}$, for some $C_1, C_2 > 0$, see (Mallat, 2012, page 1336), it follows that $\|\psi_{(-J,0)}\|_1 < \infty$ and $\|\nabla\psi_{(-J,0)}\|_1 < \infty$ (Grafakos, 2008, Chapter 2.2), and thus $\|\psi_{(-J,0)}\|_1 + \|\nabla\psi_{(-J,0)}\|_1 < \infty$. By (3.19) the output-generating atoms $\chi_n = \psi_{(-J,0)}$, $n \in \mathbb{N}_0$, satisfy the decay condition (3.17), so that all the conditions required by Theorem 1 and Corollary 1 in Section 3.3, as well as by Theorem 2 in Section 3.4.1 are satisfied.

3.5.2. Horizontal vs. vertical translation invariance

Mallat's horizontal translation invariance result (3.3),

$$\lim_{J\to\infty} |||\Phi_W(T_t f) - \Phi_W(f)||| \\ = \lim_{J\to\infty} \Big(\sum_{n=0}^{\infty} |||\Phi_W^n(T_t f) - \Phi_W^n(f)|||^2 \Big)^{1/2} = 0,$$

is asymptotic in the wavelet scale parameter J, and guarantees translation invariance in every network layer in the sense of

$$\lim_{J\to\infty} |||\Phi_W^n(T_t f) - \Phi_W^n(f)||| = 0, \qquad \forall f \in L^2(\mathbb{R}^d),\, \forall t \in \mathbb{R}^d,\, \forall n \in \mathbb{N}_0.$$

In contrast, our vertical translation invariance result (3.20) is asymptotic in the network depth n and is in line with observations made in the deep learning literature, e.g., in (Serre et al., 2005; Huang and LeCun, 2006; Mutch and Lowe, 2006; Ranzato et al., 2007; Jarrett et al., 2009), where it is found that the network's output generated at deeper layers tends to be more translation-invariant.

We can easily render Mallat's feature extractor Φ_W vertically translation-invariant according to

$$\lim_{n\to\infty} |||\Phi_W^n(T_t f) - \Phi_W^n(f)||| = 0, \qquad \forall f \in L^2(\mathbb{R}^d),\, \forall t \in \mathbb{R}^d,$$

by employing pooling by sub-sampling (i.e., $P_n = \mathrm{Id}$, $n \in \mathbb{N}$) and choosing the pooling factors such that $\lim_{n\to\infty} S_1 \cdot \ldots \cdot S_n = \infty$, see Theorem 1.

3.5.3. Deformation stability vs. sensitivity

The deformation stability bound (3.4) for scattering networks reported in (Mallat, 2012, Theorem 2.12) applies to the space

$$H_W := \Big\{ f \in L^2(\mathbb{R}^d) \;\Big|\; \|f\|_{H_W} < \infty \Big\}, \tag{3.31}$$

where

$$\|f\|_{H_W} := \sum_{n=0}^{\infty} \left(\sum_{q\in(\Lambda_W)^n} \|U[q]f\|_2^2 \right)^{1/2}.$$

Here, $(\Lambda_W)^n$ denotes the set of paths $q = \left(\lambda^{(j)}, \ldots, \lambda^{(p)}\right)$ of length n with $\lambda^{(j)}, \ldots, \lambda^{(p)} \in \Lambda_W$, see Section 3.1. While (Mallat, 2012, page 1350) cites numerical evidence on the series $\sum_{q \in (\Lambda_W)^n} \|U[q]f\|_2^2$ being finite (for some $n \in \mathbb{N}$) for a large class of signals $f \in L^2(\mathbb{R}^d)$, it seems difficult to establish this analytically, let alone to show that $\|f\|_{H_W} < \infty$. In contrast, our deformation sensitivity bound (3.23) applies *provably* to signal classes with inherent deformation insensitivity (such as, e.g. band-limited functions, cartoon functions, and Lipschitz functions). Moreover, the space H_W in (3.31) depends on the wavelet frame atoms $\{\psi_\lambda\}_{\lambda \in \Lambda_W}$, and thereby on the underlying signal transform, whereas $L_L^2(\mathbb{R}^d)$, $\mathcal{C}_{\mathrm{CART}}^K$, and V_R are, of course, completely independent of the module-sequence Ω.

Finally, Mallat's deformation stability bound (3.4) depends on the scale parameter J. This is problematic as Mallat's horizontal translation invariance result (3.3) requires $J \to \infty$, which, by $J\|D\tau\|_\infty \to \infty$ for $J \to \infty$, renders the deformation stability upper bound (3.4) void as it goes to ∞. In contrast, in our framework, the deformation sensitivity bound and the conditions for vertical translation invariance are completely decoupled.

3.5.4. Proof techniques

The techniques used in (Mallat, 2012) to prove the horizontal translation invariance result (3.3) and the deformation stability bound (3.4) make heavy use of structural specifics of the wavelet transform, namely, isotropic scaling (see, e.g., (Mallat, 2012, Appendix A)), a constant number $K \in \mathbb{N}$ of directional wavelets across scales (see, e.g., (Mallat, 2012, Equation E.1)), and several technical conditions such as a vanishing moment condition on the mother wavelet ψ (see, e.g., (Mallat, 2012, page 1391)). In addition, Mallat imposes the scattering admissibility condition (Mallat, 2012, Theorem 2.6). First of all, this condition depends on the underlying signal transform, more precisely on the mother wavelet ψ, whereas our weak admissibility condition (3.14) is in terms of the frame upper bounds B_n and the Lipschitz constants L_n and R_n. As the frame upper bounds B_n can be adjusted by

simply normalizing the frame elements, and this normalization affects neither vertical translation invariance nor deformation insensitivity, we can argue that our weak admissibility condition is independent of the signal transforms underlying the network. Second, Mallat's scattering admissibility condition plays a critical role in the proof of the horizontal translation invariance result (3.3) (see, e.g., (Mallat, 2012, page 1347)), as well as in the proof of the deformation stability bound (3.4) (see, e.g., (Mallat, 2012, Equation 2.51)). It is therefore unclear how Mallat's proof techniques could be generalized to arbitrary convolutional transforms. Third, to the best of our knowledge, no mother wavelet $\psi \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$, for $d \geqslant 2$, satisfying the scattering admissibility condition (Mallat, 2012, Theorem 2.6) has been reported in the literature. In contrast, our proof techniques are completely detached from the algebraic structures of the frames Ψ_n in the module-sequence $\Omega = \big((\Psi_n, M_n, P_n)\big)_{n\in\mathbb{N}}$. Rather, it suffices to employ (i) a module-sequence Ω that satisfies the weak admissibility condition (3.14), (ii) non-linearities M_n and operators P_n that commute with the translation operator T_t, (iii) output-generating atoms χ_n that satisfy the decay condition (3.17), and (iv) pooling factors S_n such that $\lim_{n\to\infty} S_1 \cdot S_2 \cdot \ldots \cdot S_n = \infty$. All these conditions were shown above to be easily satisfied in practice.

3.6. PROOFS

3.6.1. Proof of Proposition 3

We need to show that $\Phi_\Omega(f) \in (L^2(\mathbb{R}^d))^{\mathcal{Q}}$, for all $f \in L^2(\mathbb{R}^d)$. This will be accomplished by proving an even stronger result, namely

$$|||\Phi_\Omega(f)||| \leqslant \|f\|_2, \qquad \forall f \in L^2(\mathbb{R}^d), \tag{3.32}$$

which, by $\|f\|_2 < \infty$, establishes the claim. For ease of notation, we let $f_q := U[q]f$, for $f \in L^2(\mathbb{R}^d)$, in the following. Thanks to (3.11) and (3.14), we have $\|f_q\|_2 \leqslant \|f\|_2 < \infty$, and thus $f_q \in L^2(\mathbb{R}^d)$. The key idea of the proof is now—similarly to the proof of (Mallat, 2012,

Proposition 2.5)—to judiciously employ a telescoping series argument. We start by writing

$$|||\Phi_\Omega(f)|||^2 = \sum_{n=0}^{\infty} \sum_{q\in\Lambda^n} ||f_q * \chi_n||_2^2$$
$$= \lim_{N\to\infty} \sum_{n=0}^{N} \underbrace{\sum_{q\in\Lambda^n} ||f_q * \chi_n||_2^2}_{:=a_n}. \tag{3.33}$$

The key step is then to establish that a_n can be upper-bounded according to

$$a_n \leqslant b_n - b_{n+1}, \qquad \forall n \in \mathbb{N}_0, \tag{3.34}$$

with $b_n := \sum_{q\in\Lambda^n} \|f_q\|_2^2$, $n \in \mathbb{N}_0$, and to use this result in a telescoping series argument according to

$$\sum_{n=0}^{N} a_n \leqslant \sum_{n=0}^{N} (b_n - b_{n+1}) = (b_0 - b_1) + \cdots + (b_N - b_{N+1})$$
$$= b_0 - \underbrace{b_{N+1}}_{\geqslant 0} \leqslant b_0 = \sum_{q\in\Lambda^0} \|f_q\|_2^2 = \|U[e]f\|_2^2 = \|f\|_2^2. \tag{3.35}$$

By (3.33) this then implies (3.32). We start by noting that (3.34) reads

$$\sum_{q\in\Lambda^n} \|f_q * \chi_n\|_2^2 \leqslant \sum_{q\in\Lambda^n} ||f_q||_2^2 - \sum_{q\in\Lambda^{n+1}} \|f_q\|_2^2, \qquad \forall n \in \mathbb{N}_0, \tag{3.36}$$

and proceed by examining the second term on the RHS of (3.36). Every path

$$\tilde{q} \in \Lambda^{n+1} = \underbrace{\Lambda_1 \times \cdots \times \Lambda_n}_{=\Lambda^n} \times \Lambda_{n+1}$$

of length $n+1$ can be decomposed into a path $q \in \Lambda^n$ of length n and an index $\lambda_{n+1} \in \Lambda_{n+1}$ according to $\tilde{q} = (q, \lambda_{n+1})$. Thanks to (3.10) we have $U[\tilde{q}] = U[(q, \lambda_{n+1})] = U_{n+1}[\lambda_{n+1}]U[q]$, which yields

$$\sum_{\tilde{q}\in\Lambda^{n+1}} \|f_{\tilde{q}}\|_2^2 = \sum_{q\in\Lambda^n} \sum_{\lambda_{n+1}\in\Lambda_{n+1}} \|U_{n+1}[\lambda_{n+1}]f_q\|_2^2. \tag{3.37}$$

Substituting the second term on the RHS of (3.36) by (3.37) now yields

$$\sum_{q\in\Lambda^n} \|f_q * \chi_n\|_2^2 \leqslant \sum_{q\in\Lambda^n} \Big(||f_q||_2^2 - \sum_{\lambda_{n+1}\in\Lambda_{n+1}} \|U_{n+1}[\lambda_{n+1}] f_q\|_2^2 \Big),$$

for all $n \in \mathbb{N}_0$, which can be rewritten as

$$\begin{aligned} &\sum_{q\in\Lambda^n} \Big(\|f_q * \chi_n\|_2^2 + \sum_{\lambda_{n+1}\in\Lambda_{n+1}} \|U_{n+1}[\lambda_{n+1}] f_q\|_2^2 \Big) \\ &\leqslant \sum_{q\in\Lambda^n} ||f_q||_2^2, \qquad \forall n \in \mathbb{N}_0. \end{aligned} \tag{3.38}$$

Next, note that the second term inside the sum on the left hand side (LHS) of (3.38) can be written as

$$\begin{aligned} &\sum_{\lambda_{n+1}\in\Lambda_{n+1}} \|U_{n+1}[\lambda_{n+1}] f_q\|_2^2 = \sum_{\lambda_{n+1}\in\Lambda_{n+1}} \int_{\mathbb{R}^d} |(U_{n+1}[\lambda_{n+1}] f_q)(x)|^2 \mathrm{d}x \\ &= \sum_{\lambda_{n+1}\in\Lambda_{n+1}} S_{n+1}^d \int_{\mathbb{R}^d} \Big| P_{n+1}\big(M_{n+1}(f_q * g_{\lambda_{n+1}})\big)(S_{n+1}x) \Big|^2 \mathrm{d}x \\ &= \sum_{\lambda_{n+1}\in\Lambda_{n+1}} \int_{\mathbb{R}^d} \Big| P_{n+1}\big(M_{n+1}(f_q * g_{\lambda_{n+1}})\big)(y) \Big|^2 \mathrm{d}y \\ &= \sum_{\lambda_{n+1}\in\Lambda_{n+1}} \|P_{n+1}\big(M_{n+1}(f_q * g_{\lambda_{n+1}})\big)\|_2^2, \qquad \forall n \in \mathbb{N}_0. \end{aligned} \tag{3.39}$$

Noting that $f_q \in L^2(\mathbb{R}^d)$, as established above, and $g_{\lambda_{n+1}} \in L^1(\mathbb{R}^d)$, by assumption, it follows that $(f_q * g_{\lambda_{n+1}}) \in L^2(\mathbb{R}^d)$ thanks to Young's inequality (Grafakos, 2008, Theorem 1.2.12). We use the Lipschitz property of M_{n+1} and P_{n+1}, i.e., $\|M_{n+1}(f_q * g_{\lambda_{n+1}}) - M_{n+1}h\|_2 \leqslant L_{n+1}\|f_q * g_{\lambda_{n+1}} - h\|$, and $\|P_{n+1}(f_q * g_{\lambda_{n+1}}) - P_{n+1}h\|_2 \leqslant R_{n+1}\|f_q * g_{\lambda_{n+1}} - h\|$, together with $M_{n+1}h = 0$ and $P_{n+1}h = 0$ for $h = 0$, to upper-bound the term inside the sum in (3.39) according to

$$\begin{aligned} \|P_{n+1}\big(M_{n+1}(f_q * g_{\lambda_{n+1}})\big)\|_2^2 &\leqslant R_{n+1}^2 \|M_{n+1}(f_q * g_{\lambda_{n+1}})\|_2^2 \\ &\leqslant L_{n+1}^2 R_{n+1}^2 \|f_q * g_{\lambda_{n+1}}\|_2^2, \end{aligned} \tag{3.40}$$

for all $n \in \mathbb{N}_0$. Substituting the second term inside the sum on the LHS of (3.38) by the upper bound resulting from insertion of (3.40) into (3.39) yields

$$\begin{aligned}
&\sum_{q\in\Lambda^n} \Big(\|f_q * \chi_n\|_2^2 + L_{n+1}^2 R_{n+1}^2 \sum_{\lambda_{n+1}\in\Lambda_{n+1}} \|f_q * g_{\lambda_{n+1}}\|_2^2 \Big) \\
&\leqslant \sum_{q\in\Lambda^n} \max\{1, L_{n+1}^2 R_{n+1}^2\} \Big(\|f_q * \chi_n\|_2^2 \\
&+ \sum_{\lambda_{n+1}\in\Lambda_{n+1}} \|f_q * g_{\lambda_{n+1}}\|_2^2 \Big), \qquad \forall n \in \mathbb{N}_0. \qquad (3.41)
\end{aligned}$$

As the functions $\{g_{\lambda_{n+1}}\}_{\lambda_{n+1}\in\Lambda_{n+1}} \cup \{\chi_n\}$ are the atoms of the semi-discrete frame Ψ_{n+1} for $L^2(\mathbb{R}^d)$ and $f_q \in L^2(\mathbb{R}^d)$, as established above, we have

$$\|f_q * \chi_n\|_2^2 + \sum_{\lambda_{n+1}\in\Lambda_{n+1}} \|f_q * g_{\lambda_{n+1}}\|_2^2 \leqslant B_{n+1}\|f_q\|_2^2,$$

which, when used in (3.41) yields

$$\begin{aligned}
&\sum_{q\in\Lambda^n} \Big(\|f_q * \chi_n\|_2^2 + \sum_{\lambda_{n+1}\in\Lambda_{n+1}} \|U_{n+1}[\lambda_{n+1}] f_q\|_2^2 \Big) \\
&\leqslant \sum_{q\in\Lambda^n} \max\{1, L_{n+1}^2 R_{n+1}^2\} B_{n+1} \|f_q\|_2^2 \\
&= \sum_{q\in\Lambda^n} \max\{B_{n+1}, B_{n+1} L_{n+1}^2 R_{n+1}^2\} \|f_q\|_2^2, \qquad \forall n \in \mathbb{N}_0. \qquad (3.42)
\end{aligned}$$

Finally, invoking the assumption

$$\max\{B_n, B_n L_n^2 R_{n+1}^2\} \leqslant 1, \qquad \forall n \in \mathbb{N},$$

in (3.42) yields (3.38) and thereby completes the proof.

3.6.2. Proof of Theorem 1

We start by proving i). The key step in establishing (3.16) is to show that the operator U_n, $n \in \mathbb{N}$, defined in (3.7) satisfies the relation

$$U_n[\lambda_n] T_t f = T_{t/S_n} U_n[\lambda_n] f, \qquad (3.43)$$

for all $f \in L^2(\mathbb{R}^d)$, all $t \in \mathbb{R}^d$, and all $\lambda_n \in \Lambda_n$. With the definition of $U[q]$ in (3.10) this then yields

$$U[q]T_t f = T_{t/(S_1 \cdots S_n)} U[q] f, \tag{3.44}$$

for all $f \in L^2(\mathbb{R}^d)$, all $t \in \mathbb{R}^d$, and all $\lambda_n \in \Lambda_n$. The identity (3.16) is then a direct consequence of (3.44) and the translation-covariance of the convolution operator:

$$\begin{aligned}
\Phi_\Omega^n(T_t f) &= \big\{ \big(U[q]T_t f\big) * \chi_n \big\}_{q \in \Lambda^n} = \big\{ \big(T_{t/(S_1 \cdots S_n)} U[q] f\big) * \chi_n \big\}_{q \in \Lambda^n} \\
&= \big\{ T_{t/(S_1 \cdots S_n)} \big((U[q]f) * \chi_n \big) \big\}_{q \in \Lambda^n} \\
&= T_{t/(S_1 \cdots S_n)} \big\{ (U[q]f) * \chi_n \big\}_{q \in \Lambda^n} \\
&= T_{t/(S_1 \cdots S_n)} \Phi_\Omega^n(f), \qquad \forall f \in L^2(\mathbb{R}^d),\, \forall t \in \mathbb{R}^d.
\end{aligned}$$

To establish (3.43), we first define the unitary operator $D_n : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$, $D_n f := S_n^{d/2} f(S_n \cdot)$, and note that

$$\begin{aligned}
U_n[\lambda_n] T_t f &= S_n^{d/2} P_n \Big(M_n \big((T_t f) * g_{\lambda_n} \big) \Big) (S_n \cdot) \\
&= D_n P_n \Big(M_n \big((T_t f) * g_{\lambda_n} \big) \Big) \\
&= D_n P_n \Big(M_n \big(T_t (f * g_{\lambda_n}) \big) \Big) \\
&= D_n P_n \Big(T_t \big(M_n (f * g_{\lambda_n}) \big) \Big) \qquad (3.45) \\
&= D_n T_t \Big(P_n \Big(\big(M_n (f * g_{\lambda_n}) \big) \Big) \Big), \qquad (3.46)
\end{aligned}$$

for all $f \in L^2(\mathbb{R}^d)$ and all $t \in \mathbb{R}^d$, where in (3.45) and (3.46) we employed $M_n T_t = T_t M_n$ and $P_n T_t = T_t P_n$, for all $n \in \mathbb{N}$, and all $t \in \mathbb{R}^d$, respectively, both of which are by assumption. Next, using

$$D_n T_t f = S_n^{d/2} f(S_n \cdot - t) = S_n^{d/2} f(S_n(\cdot - t/S_n)) = T_{t/S_n} D_n f,$$

for all $f \in L^2(\mathbb{R}^d)$ and all $t \in \mathbb{R}^d$ in (3.46) yields

$$\begin{aligned}
U_n[\lambda_n] T_t f &= D_n T_t \Big(P_n \Big(\big(M_n (f * g_{\lambda_n}) \big) \Big) \Big) \\
&= T_{t/S_n} \Big(D_n P_n \Big(\big(M_n (f * g_{\lambda_n}) \big) \Big) \Big) = T_{t/S_n} U_n[\lambda_n] f,
\end{aligned}$$

for all $f \in L^2(\mathbb{R}^d)$ and all $t \in \mathbb{R}^d$. This completes the proof of i).

Next, we prove ii). For ease of notation, again, we let $f_q := U[q]f$, for $f \in L^2(\mathbb{R}^d)$. Thanks to (3.11) and the admissibility condition (3.14), we have $\|f_q\|_2 \leqslant \|f\|_2 < \infty$, and thus $f_q \in L^2(\mathbb{R}^d)$. We first write

$$|||\Phi_\Omega^n(T_t f) - \Phi_\Omega^n(f)|||^2 = |||T_{t/(S_1 \cdots S_n)}\Phi_\Omega^n(f) - \Phi_\Omega^n(f)|||^2 \tag{3.47}$$

$$= \sum_{q \in \Lambda^n} \|T_{t/(S_1 \cdots S_n)}(f_q * \chi_n) - f_q * \chi_n\|_2^2$$

$$= \sum_{q \in \Lambda^n} \|M_{-t/(S_1 \cdots S_n)}(\widehat{f_q * \chi_n}) - \widehat{f_q * \chi_n}\|_2^2, \tag{3.48}$$

for all $n \in \mathbb{N}$, where in (3.47) we used (3.16), and in (3.48) we employed Parseval's formula (Rudin, 1991, page 189)—noting that $(f_q * \chi_n) \in L^2(\mathbb{R}^d)$ thanks to Young's inequality (Grafakos, 2008, Theorem 1.2.12)—together with the relation $\widehat{T_t f} = M_{-t}\widehat{f}$, for all $f \in L^2(\mathbb{R}^d)$ and all $t \in \mathbb{R}^d$. The key step is then to establish the upper bound

$$\|M_{-t/(S_1 \cdots S_n)}(\widehat{f_q * \chi_n}) - \widehat{f_q * \chi_n}\|_2^2 \leqslant \frac{4\pi^2|t|^2K^2}{(S_1 \cdots S_n)^2}\|f_q\|_2^2, \tag{3.49}$$

for all $n \in \mathbb{N}$, where $K > 0$ corresponds to the constant in the decay condition (3.17), and to note that

$$\sum_{q \in \Lambda^n} \|f_q\|_2^2 \leqslant \sum_{q \in \Lambda^{n-1}} \|f_q\|_2^2, \qquad \forall n \in \mathbb{N}, \tag{3.50}$$

which follows from (3.34) thanks to

$$0 \leqslant \sum_{q \in \Lambda^{n-1}} ||f_q * \chi_{n-1}||_2^2 = a_{n-1}$$

$$\leqslant b_{n-1} - b_n = \sum_{q \in \Lambda^{n-1}} \|f_q\|_2^2 - \sum_{q \in \Lambda^n} \|f_q\|_2^2, \qquad \forall n \in \mathbb{N}.$$

Iterating on (3.50) yields

$$\sum_{q \in \Lambda^n} \|f_q\|_2^2 \leqslant \sum_{q \in \Lambda^{n-1}} \|f_q\|_2^2 \leqslant \ldots \leqslant \sum_{q \in \Lambda^0} \|f_q\|_2^2$$

$$= \|U[e]f\|_2^2 = \|f\|_2^2, \qquad \forall n \in \mathbb{N}. \tag{3.51}$$

The identity (3.48) together with the inequalities (3.49) and (3.51) then directly imply

$$|||\Phi_\Omega^n(T_t f) - \Phi_\Omega^n(f)|||^2 \leqslant \frac{4\pi^2|t|^2K^2}{(S_1\cdots S_n)^2}\|f\|_2^2, \qquad \forall n \in \mathbb{N}. \tag{3.52}$$

It remains to prove (3.49). To this end, we first note that

$$\begin{aligned}&\|M_{-t/(S_1\cdots S_n)}(\widehat{f_q * \chi_n}) - \widehat{f_q * \chi_n}\|_2^2 \\ &= \int_{\mathbb{R}^d} \left|e^{-2\pi i\langle t,\omega\rangle/(S_1\cdots S_n)} - 1\right|^2 |\widehat{\chi_n}(\omega)|^2 |\widehat{f}_q(\omega)|^2 \mathrm{d}\omega. \end{aligned}\tag{3.53}$$

Since $|e^{-2\pi i x} - 1| \leqslant 2\pi|x|$, for all $x \in \mathbb{R}$, it follows that

$$|e^{-2\pi i\langle t,\omega\rangle/(S_1\cdots S_n)} - 1|^2 \leqslant \frac{4\pi^2|\langle t,\omega\rangle|^2}{(S_1\cdots S_n)^2} \leqslant \frac{4\pi^2|t|^2|\omega|^2}{(S_1\cdots S_n)^2}, \tag{3.54}$$

where in the last step we employed the Cauchy-Schwartz inequality. Substituting (3.54) into (3.53) yields

$$\begin{aligned}&\|M_{-t/(S_1\cdots S_n)}(\widehat{f_q * \chi_n}) - \widehat{f_q * \chi_n}\|_2^2 \\ &\leqslant \frac{4\pi^2|t|^2}{(S_1\cdots S_n)^2} \int_{\mathbb{R}^d} |\omega|^2 |\widehat{\chi_n}(\omega)|^2 |\widehat{f}_q(\omega)|^2 \mathrm{d}\omega \\ &\leqslant \frac{4\pi^2|t|^2K^2}{(S_1\cdots S_n)^2} \int_{\mathbb{R}^d} |\widehat{f}_q(\omega)|^2 \mathrm{d}\omega \qquad (3.55) \\ &= \frac{4\pi^2|t|^2K^2}{(S_1\cdots S_n)^2} \|\widehat{f}_q\|_2^2 = \frac{4\pi^2|t|^2K^2}{(S_1\cdots S_n)^2} \|f_q\|_2^2, \qquad \forall n \in \mathbb{N}, \qquad (3.56)\end{aligned}$$

where in (3.55) we employed the decay condition (3.17), and in the last step, again, we used Parseval's formula (Rudin, 1991, page 189). This establishes (3.49) and thereby completes the proof of ii).

3.6.3. Proof of Corollary 1

The key idea of the proof is—similarly to the proof of ii) in Theorem 1 in Section 3.6.2—to upper-bound the deviation from perfect covariance in the frequency domain. For ease of notation, again, we let

$f_q := U[q]f$, for $f \in L^2(\mathbb{R}^d)$. Thanks to (3.11) and the admissibility condition (3.14), we have $\|f_q\|_2 \leqslant \|f\|_2 < \infty$, and thus $f_q \in L^2(\mathbb{R}^d)$. We first write

$$\begin{aligned} |||\Phi_\Omega^n(T_t f) - T_t \Phi_\Omega^n(f)|||^2 &= |||T_{t/(S_1 \cdots S_n)} \Phi_\Omega^n(f) - T_t \Phi_\Omega^n(f)|||^2 \quad (3.57) \\ &= \sum_{q \in \Lambda_1^n} \|(T_{t/(S_1 \cdots S_n)} - T_t)(f_q * \chi_n)\|_2^2 \\ &= \sum_{q \in \Lambda_1^n} \|(M_{-t/(S_1 \cdots S_n)} - M_{-t})\widehat{(f_q * \chi_n)}\|_2^2, \quad (3.58) \end{aligned}$$

for all $n \in \mathbb{N}$, where in (3.57) we used (3.16), and in (3.58) we employed Parseval's formula (Rudin, 1991, page 189)—noting that $(f_q * \chi_n) \in L^2(\mathbb{R}^d)$ thanks to Young's inequality (Grafakos, 2008, Theorem 1.2.12)—together with the relation $\widehat{T_t f} = M_{-t}\widehat{f}$, for all $f \in L^2(\mathbb{R}^d)$, and all $t \in \mathbb{R}^d$. The key step is then to establish the upper bound

$$\begin{aligned} &\|(M_{-t/(S_1 \cdots S_n)} - M_{-t})\widehat{(f_q * \chi_n)}\|_2^2 \\ &\leqslant 4\pi^2 |t|^2 K^2 \big|1/(S_1 \cdots S_n) - 1\big|^2 \|f_q\|_2^2, \quad (3.59) \end{aligned}$$

where $K > 0$ corresponds to the constant in the decay condition (3.17). Arguments similar to those leading to (3.52) then complete the proof. It remains to prove (3.59):

$$\begin{aligned} &\|(M_{-t/(S_1 \cdots S_n)} - M_{-t})\widehat{(f_q * \chi_n)}\|_2^2 \\ &= \int_{\mathbb{R}^d} \big|e^{-2\pi i \langle t, \omega \rangle / (S_1 \cdots S_n)} - e^{-2\pi i \langle t, \omega \rangle}\big|^2 |\widehat{\chi_n}(\omega)|^2 |\widehat{f_q}(\omega)|^2 \mathrm{d}\omega. \quad (3.60) \end{aligned}$$

Since $|e^{-2\pi i x} - e^{-2\pi i y}| \leqslant 2\pi |x - y|$, for all $x, y \in \mathbb{R}$, it follows that

$$\begin{aligned} &\big|e^{-2\pi i \langle t, \omega \rangle / (S_1 \cdots S_n)} - e^{-2\pi i \langle t, \omega \rangle}\big|^2 \\ &\leqslant 4\pi^2 |t|^2 |\omega|^2 \big|1/(S_1 \cdots S_n) - 1\big|^2, \quad (3.61) \end{aligned}$$

where, again, we employed the Cauchy-Schwartz inequality. Substituting (3.61) into (3.60), and employing arguments similar to those leading to (3.56), establishes (3.59) and thereby completes the proof.

3.6.4. Proof of Theorem 2

As already mentioned at the beginning of Section 3.4, the proof of the deformation sensitivity bound (3.23) is based on two key ingredients. The first one, stated in Proposition 7 in Section 3.6.8, establishes that the feature extractor Φ_Ω is Lipschitz-continuous with Lipschitz constant $L_\Omega = 1$, i.e.,

$$|||\Phi_\Omega(f) - \Phi_\Omega(h)||| \leqslant \|f - h\|_2, \qquad \forall f, h \in L^2(\mathbb{R}^d), \tag{3.62}$$

and needs the admissibility condition (3.14) only. The second ingredient is an upper bound on the deformation error $\|f - F_\tau f\|_2$ according to (see Definition 6 in Section 3.4.1)

$$\|f - F_\tau f\|_2 \leqslant C\|\tau\|_\infty^\alpha \|f\|_2, \tag{3.63}$$

and is established in Proposition 4 in Section 3.4.2 for band-limited functions, in Proposition 5 in Section 3.4.3 for cartoon functions, and in Proposition 6 in Section 3.4.4 for Lipschitz functions. We now show how (3.62) and (3.63) can be combined to establish the deformation sensitivity bound (3.23). To this end, we first apply (3.62) with $h := F_\tau f = f(\cdot - \tau(\cdot))$ to get

$$|||\Phi_\Omega(f) - \Phi_\Omega(F_\tau f)||| \leqslant \|f - F_\tau f\|_2, \qquad \forall f \in L^2(\mathbb{R}^d). \tag{3.64}$$

Here, we used $F_\tau f \in L^2(\mathbb{R}^d)$, which is thanks to

$$\|F_\tau f\|_2^2 = \int_{\mathbb{R}^d} |f(x - \tau(x))|^2 \mathrm{d}x \leqslant 2\|f\|_2^2,$$

obtained through the change of variables $u = x - \tau(x)$, together with

$$\frac{\mathrm{d}u}{\mathrm{d}x} = |\det(E - (D\tau)(x))| \geqslant 1 - d\|D\tau\|_\infty \geqslant 1/2, \qquad \forall x \in \mathbb{R}^d. \tag{3.65}$$

The first inequality in (3.65) follows from:

Lemma 2. *(Brent et al., 2015, Corollary 1) Let $M \in \mathbb{R}^{d\times d}$ be such that $|M_{i,j}| \leqslant \alpha$, for all i, j with $1 \leqslant i, j \leqslant d$. If $d\alpha \leqslant 1$, then*

$$|\det(E - M)| \geqslant 1 - d\alpha.$$

The second inequality in (3.65) is a consequence of the assumption $\|D\tau\|_\infty \leqslant \frac{1}{2d}$. The proof is finalized by replacing the RHS of (3.64) by the RHS of (3.63).

3.6.5. Proof of Proposition 4

We first determine an integral operator

$$(Kf)(x) = \int_{\mathbb{R}^d} k(x,u) f(u) \mathrm{d}u \tag{3.66}$$

satisfying the signal-class specific identity $Kf = F_\tau f - f$, for all $f \in L^2_L(\mathbb{R}^d)$, and then upper-bound the deformation error $\|f - F_\tau f\|_2$ according to

$$\|f - F_\tau f\|_2 = \|F_\tau f - f\|_2 = \|Kf\|_2 \leqslant \|K\|_{2,2} \|f\|_2,$$

for all $f \in L^2_L(\mathbb{R}^d)$. Application of Schur's Lemma, stated below, then yields

$$\|K\|_{2,2} \leqslant CL\|\tau\|_\infty,$$

for some $C > 0$, which completes the proof.

Schur's Lemma. *(Grafakos, 2008, Appendix I.1) Let $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{C}$ be a locally integrable function satisfying*

$$(i) \ \sup_{x \in \mathbb{R}^d} \int_{\mathbb{R}^d} |k(x,u)| \mathrm{d}u \leqslant \alpha, \quad (ii) \ \sup_{u \in \mathbb{R}^d} \int_{\mathbb{R}^d} |k(x,u)| \mathrm{d}x \leqslant \alpha, \tag{3.67}$$

where $\alpha > 0$. Then, $(Kf)(x) = \int_{\mathbb{R}^d} k(x,u) f(u) \mathrm{d}u$ is a bounded operator from $L^2(\mathbb{R}^d)$ to $L^2(\mathbb{R}^d)$ with operator norm $\|K\|_{2,2} \leqslant \alpha$.

We start by determining the integral operator K in (3.66). To this end, consider $\eta \in S(\mathbb{R}^d, \mathbb{C})$ such that $\widehat{\eta}(\omega) = 1$, for all $\omega \in B_1(0)$. Setting $\gamma(x) := L^d \eta(Lx)$ yields $\gamma \in S(\mathbb{R}^d, \mathbb{C})$ and $\widehat{\gamma}(\omega) = \widehat{\eta}(\omega/L)$. Thus, $\widehat{\gamma}(\omega) = 1$, for all $\omega \in B_L(0)$, and hence $\widehat{f} = \widehat{f} \cdot \widehat{\gamma}$, so that $f = f * \gamma$, for all $f \in L^2_L(\mathbb{R}^d)$. Next, we define the operator $A_\gamma : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$, $A_\gamma f := f * \gamma$, and note that A_γ is well-defined, i.e., $A_\gamma f \in L^2(\mathbb{R}^d)$, for all $f \in L^2(\mathbb{R}^d)$, thanks to Young's

inequality (Grafakos, 2008, Theorem 1.2.12) (since $f \in L^2(\mathbb{R}^d)$ and $\gamma \in S(\mathbb{R}^d, \mathbb{C}) \subseteq L^1(\mathbb{R}^d)$). Moreover, $A_\gamma f = f$, for all $f \in L^2_L(\mathbb{R}^d)$. Setting $K := F_\tau A_\gamma - A_\gamma$, we get $Kf = F_\tau A_\gamma f - A_\gamma f = F_\tau f - f$, for all $f \in L^2_L(\mathbb{R}^d)$, as desired. Furthermore, it follows from

$$(F_\tau A_\gamma f)(x) = \int_{\mathbb{R}^d} \gamma(x - \tau(x) - u) f(u) \mathrm{d}u,$$

that the integral operator $K = F_\tau A_\gamma - A_\gamma$, i.e., $(Kf)(x) = \int_{\mathbb{R}^d} k(x,u) f(u) \mathrm{d}u$, has the kernel

$$k(x,u) := \gamma(x - \tau(x) - u) - \gamma(x - u). \tag{3.68}$$

Before we can apply Schur's Lemma to establish an upper bound on $\|K\|_{2,2}$, we need to verify that k in (3.68) is locally integrable, i.e., we need to show that for every compact set $S \subseteq \mathbb{R}^d \times \mathbb{R}^d$ we have $\int_S |k(x,u)| \mathrm{d}(x,u) < \infty$. To this end, let $S \subseteq \mathbb{R}^d \times \mathbb{R}^d$ be a compact set. Next, choose compact sets $S_1, S_2 \subseteq \mathbb{R}^d$ such that $S \subseteq S_1 \times S_2$. Thanks to $\gamma \in S(\mathbb{R}^d, \mathbb{C})$, $\tau \in C^1(\mathbb{R}^d, \mathbb{R}^d)$, and $\omega \in C(\mathbb{R}^d, \mathbb{R})$, all by assumption, the function $|k| : S_1 \times S_2 \to \mathbb{C}$ is continuous as a composition of continuous functions, and therefore also Lebesgue-measurable. We further have

$$\begin{aligned}
&\int_{S_1} \int_{S_2} |k(x,u)| \mathrm{d}x \mathrm{d}u \leqslant \int_{S_1} \int_{\mathbb{R}^d} |k(x,u)| \mathrm{d}x \mathrm{d}u \\
&\leqslant \int_{S_1} \int_{\mathbb{R}^d} |\gamma(x - \tau(x) - u)| \mathrm{d}x \mathrm{d}u + \int_{S_1} \int_{\mathbb{R}^d} |\gamma(x-u)| \mathrm{d}x \mathrm{d}u \\
&\leqslant 2 \int_{S_1} \int_{\mathbb{R}^d} |\gamma(y)| \mathrm{d}y \mathrm{d}u + \int_{S_1} \int_{\mathbb{R}^d} |\gamma(y)| \mathrm{d}y \, \mathrm{d}u \qquad (3.69) \\
&= 3\mu_L(S_1) \|\gamma\|_1 < \infty, \qquad (3.70)
\end{aligned}$$

where the first term in (3.69) follows by the change of variables $y = x - \tau(x) - u$, together with

$$\frac{\mathrm{d}y}{\mathrm{d}x} = |\det(E - (D\tau)(x))| \geqslant 1 - d\|D\tau\|_\infty \geqslant 1/2, \qquad \forall x \in \mathbb{R}^d. \tag{3.71}$$

The arguments underlying (3.71) were already detailed at the end of Section 3.6.4. It follows that k is locally integrable owing to

$$\int_S |k(x,u)|\mathrm{d}(x,u) \leqslant \int_{S_1\times S_2} |k(x,u)|\mathrm{d}(x,u) \tag{3.72}$$

$$= \int_{S_1}\int_{S_2} |k(x,u)|\mathrm{d}x\mathrm{d}u < \infty, \tag{3.73}$$

where (3.72) follows from $S \subseteq S_1 \times S_2$, (3.73) is thanks to the Fubini-Tonelli Theorem (DiBenedetto, 2002, Theorem 14.2) noting that $|k| : S_1 \times S_2 \to \mathbb{C}$ is Lebesgue-measurable (as established above) and non-negative, and the last step is due to (3.70). Next, we need to verify conditions (i) and (ii) in (3.67) and determine the corresponding $\alpha > 0$. In fact, we seek a specific constant α of the form

$$\alpha = CL\|\tau\|_\infty, \tag{3.74}$$

for some $C > 0$. This will be accomplished as follows: For $x, u \in \mathbb{R}^d$, we parametrize the integral kernel in (3.68) according to $h_{x,u}(t) := \gamma(x - t\tau(x) - u) - \gamma(x - u)$. A Taylor series expansion (Rudin, 1983, page 411) of $h_{x,u}(t)$ w.r.t. the variable t now yields

$$h_{x,u}(t) = \underbrace{h_{x,u}(0)}_{=0} + \int_0^t h'_{x,u}(\lambda)\mathrm{d}\lambda = \int_0^t h'_{x,u}(\lambda)\mathrm{d}\lambda, \tag{3.75}$$

for all $t \in \mathbb{R}$, where $h'_{x,u}(t) = (\frac{\mathrm{d}}{\mathrm{d}t}h_{x,u})(t)$. Note that $h_{x,u} \in C^1(\mathbb{R}, \mathbb{C})$ thanks to $\gamma \in S(\mathbb{R}^d, \mathbb{C})$. Setting $t = 1$ in (3.75) we get

$$|k(x,u)| = |h_{x,u}(1)| \leqslant \int_0^1 |h'_{x,u}(\lambda)|\mathrm{d}\lambda, \tag{3.76}$$

where

$$h'_{x,u}(\lambda) = -\left\langle \nabla\gamma(x - \lambda\tau(x) - u), \tau(x)\right\rangle,$$

for $\lambda \in [0,1]$. We further have

$$|h'_{x,u}(\lambda)| \leqslant \left|\left\langle \nabla\gamma(x - \lambda\tau(x) - u), \tau(x)\right\rangle\right|$$

$$\leqslant |\tau(x)||\nabla\gamma(x - \lambda\tau(x) - u)|. \tag{3.77}$$

Now, using $|\tau(x)| \leqslant \sup_{y\in\mathbb{R}^d} |\tau(y)| = \|\tau\|_\infty$ in (3.77), together with (3.76), we get the upper bound

$$|k(x,u)| \leqslant \|\tau\|_\infty \int_0^1 |\nabla\gamma(x-\lambda\tau(x)-u)|\mathrm{d}\lambda. \tag{3.78}$$

Next, we integrate (3.78) w.r.t. u to establish (i) in (3.67):

$$\begin{aligned}\int_{\mathbb{R}^d} |k(x,u)|\mathrm{d}u &\leqslant \|\tau\|_\infty \int_{\mathbb{R}^d}\int_0^1 |\nabla\gamma(x-\lambda\tau(x)-u)|\mathrm{d}\lambda\mathrm{d}u \\ &= \|\tau\|_\infty \int_0^1\int_{\mathbb{R}^d} |\nabla\gamma(x-\lambda\tau(x)-u)|\mathrm{d}u\mathrm{d}\lambda \qquad (3.79)\\ &= \|\tau\|_\infty \int_0^1\int_{\mathbb{R}^d} |\nabla\gamma(y)|\mathrm{d}y\mathrm{d}\lambda \\ &= \|\tau\|_\infty \|\nabla\gamma\|_1, \qquad (3.80)\end{aligned}$$

where (3.79) follows by application of the Fubini-Tonelli Theorem (DiBenedetto, 2002, Theorem 14.2) noting that the functions $(u,\lambda)\mapsto |\nabla\gamma(x-\lambda\tau(x)-u)|$, $(u,\lambda)\in\mathbb{R}^d\times[0,1]$, and $(u,\lambda)\mapsto|\gamma(x-\lambda\tau(x)-u)|$, $(u,\lambda)\in\mathbb{R}^d\times[0,1]$, are both non-negative and continuous (and thus Lebesgue-measurable) as compositions of continuous functions. Finally, using $\gamma = L^d\eta(L\cdot)$, and thus $\nabla\gamma = L^{d+1}\nabla\eta(L\cdot)$ and $\|\nabla\gamma\|_1 = L\|\nabla\eta\|_1$ in (3.80) yields

$$\sup_{x\in\mathbb{R}^d}\int_{\mathbb{R}^d} |k(x,u)|\mathrm{d}u \leqslant L\|\nabla\eta\|_1\|\tau\|_\infty, \tag{3.81}$$

which establishes an upper bound of the form (i) in (3.67) that exhibits the desired structure for α. Condition (ii) in (3.67) is established similarly by integrating (3.78) w.r.t. x according to

$$\begin{aligned}\int_{\mathbb{R}^d} |k(x,u)|\mathrm{d}x &\leqslant \|\tau\|_\infty \int_{\mathbb{R}^d}\int_0^1 |\nabla\gamma(x-\lambda\tau(x)-u)|\mathrm{d}\lambda\mathrm{d}x \\ &= \|\tau\|_\infty \int_0^1\int_{\mathbb{R}^d} |\nabla\gamma(x-\lambda\tau(x)-u)|\mathrm{d}x\mathrm{d}\lambda \qquad (3.82)\\ &\leqslant 2\,\|\tau\|_\infty \int_0^1\int_{\mathbb{R}^d} |\nabla\gamma(y)|\mathrm{d}y\mathrm{d}\lambda \qquad (3.83)\end{aligned}$$

$$= 2\,\|\tau\|_\infty \|\nabla\gamma\|_1 = 2\,L\,\|\nabla\eta\|_1 \|\tau\|_\infty, \tag{3.84}$$

which yields an upper bound of the form (ii) in (3.67) with the desired structure for α. Here, again, (3.82) follows by application of the Fubini-Tonelli Theorem (DiBenedetto, 2002, Theorem 14.2) noting that the functions $(x,\lambda) \mapsto |\nabla\gamma(x - \lambda\tau(x) - u)|$, $(x,\lambda) \in \mathbb{R}^d \times [0,1]$, and $(x,\lambda) \mapsto |\gamma(x - \lambda\tau(x) - u)|$, $(x,\lambda) \in \mathbb{R}^d \times [0,1]$, are both non-negative and continuous (and thus Lebesgue-measurable) as a composition of continuous functions. The inequality (3.83) follows from a change of variables argument similar to the one in (3.69) and (3.71). Combining (3.81) and (3.84), we finally get (3.74) with $C := 2\|\nabla\eta\|_1$. This completes the proof.

3.6.6. Proof of Proposition 5

The proof of (3.28) is based on judiciously combining deformation sensitivity bounds for the components f_1, f_2 in $(f_1 + \mathbb{1}_B f_2) \in \mathcal{C}^K_{\mathrm{CART}}$ and for the indicator function $\mathbb{1}_B$. The first bound, stated in Proposition 6 in Section 3.4.4, reads

$$\|f - F_\tau f\|_2 \le C_K \|\tau\|_\infty, \tag{3.85}$$

and applies to functions f satisfying the decay condition

$$|\nabla f(x)| \le K \langle x \rangle^{-d}, \tag{3.86}$$

with the constant $C_K > 0$ not depending on f, τ (see (3.93)). The bound in (3.85) needs the assumption $\|\tau\|_\infty < \frac{1}{2}$. The second bound, stated in Lemma 3 below, is

$$\|\mathbb{1}_B - F_\tau \mathbb{1}_B\|_2 \le C_{\partial B}^{1/2} \|\tau\|_\infty^{1/2}, \tag{3.87}$$

where the constant $C_{\partial B} > 0$ is independent of τ. We now show how (3.85) and (3.87) can be combined to establish (3.28). For $f = (f_1 + \mathbb{1}_B f_2) \in \mathcal{C}^K_{\mathrm{CART}}$, we have

$$\begin{aligned} \|f - F_\tau f\|_2 &\leqslant \|f_1 - F_\tau f_1\|_2 \\ &+ \|\mathbb{1}_B (f_2 - F_\tau f_2)\|_2 + \|(\mathbb{1}_B - F_\tau \mathbb{1}_B)(F_\tau f_2)\|_2 \\ &\leqslant \|f_1 - F_\tau f_1\|_2 + \|f_2 - F_\tau f_2\|_2 + \|\mathbb{1}_B - F_\tau \mathbb{1}_B\|_2 \|F_\tau f_2\|_\infty, \end{aligned} \tag{3.88}$$

where in (3.88) we used $F_\tau(\mathbb{1}_B f_2)(x) = (\mathbb{1}_B f_2)(x-\tau(x)) = \mathbb{1}_B(x-\tau(x))f_2((x-\tau(x))) = (F_\tau \mathbb{1}_B)(x)(F_\tau f_2)(x)$. With the upper bounds (3.85) and (3.87), invoking properties of the class of cartoon functions $\mathcal{C}^K_{\mathrm{CART}}$ (namely, (i) f_1,f_2 satisfy (3.86) and thus (3.85), and (ii) $\|F_\tau f_2\|_\infty = \sup_{x\in\mathbb{R}^d} |f_2(x-\tau(x))| \leqslant \sup_{y\in\mathbb{R}^d} |f_2(y)| = \|f_2\|_\infty \leqslant K$), this yields

$$\begin{aligned}\|f - F_\tau f\|_2 &\leqslant 2\, C_K \, \|\tau\|_\infty + K C_{\partial B}^{1/2} \|\tau\|_\infty^{1/2} \\ &\leqslant \underbrace{2\max\{2\, C_K, K C_{\partial B}^{1/2}\}}_{=:C'_K} \|\tau\|_\infty^{1/2},\end{aligned}$$

which completes the proof of (3.28).

We continue with the deformation sensitivity result (3.87) for indicator functions $\mathbb{1}_B$.

Lemma 3. *Let $B \subseteq \mathbb{R}^d$ be a compact domain whose boundary ∂B is a compact topologically embedded C^2-hypersurface of $\mathbb{R}^d$ without boundary. Then, there exists a constant $C_{\partial B} > 0$ (that does not depend on τ) such that for all $\tau : \mathbb{R}^d \to \mathbb{R}^d$ with $\|\tau\|_\infty \leqslant 1$, it holds that*

$$\|\mathbb{1}_B - F_\tau \mathbb{1}_B\|_2 \leq C_{\partial B}^{1/2} \|\tau\|_\infty^{1/2}.$$

Proof. In order to upper-bound

$$\|\mathbb{1}_B - F_\tau \mathbb{1}_B\|_2^2 = \int_{\mathbb{R}^d} |\mathbb{1}_B(x) - \mathbb{1}_B(x-\tau(x))|^2 \mathrm{d}x,$$

we first note that the integrand $h(x) := |\mathbb{1}_B(x) - \mathbb{1}_B(x-\tau(x))|^2$ satisfies $h(x) = 1$, for $x \in S$, where

$$\begin{aligned}S :&= \{x \in \mathbb{R}^d \,|\, x \in B \;\text{ and }\; x - \tau(x) \notin B\} \\ &\cup \{x \in \mathbb{R}^d \,|\, x \notin B \;\text{ and }\; x - \tau(x) \in B\},\end{aligned}$$

and $h(x) = 0$, for $x \in \mathbb{R}^d \backslash S$. Moreover, owing to $S \subseteq \big(\partial B + B_{\|\tau\|_\infty}(0)\big)$, where $(\partial B + B_{\|\tau\|_\infty}(0))$ is a tube of radius $\|\tau\|_\infty$ around the boundary ∂B of B, and Lemma 4, stated below, there exists a constant $C_{\partial B} > 0$ such that

$$\mathrm{vol}^d(S) \leqslant \mathrm{vol}^d(\partial B + B_{\|\tau\|_\infty}(0)) \leqslant C_{\partial B} \|\tau\|_\infty, \tag{3.89}$$

for all τ with $\|\tau\|_\infty \leqslant 1$. We therefore have

$$\|\mathbb{1}_B - F_\tau \mathbb{1}_B\|_2^2 = \int_{\mathbb{R}^d} h(x)\mathrm{d}x = \int_S 1\mathrm{d}x = \mathrm{vol}^d(S) \leqslant C_{\partial B}\|\tau\|_\infty,$$

which completes the proof. □

It remains to establish the second inequality in (3.89).

Lemma 4. *Let $M \subseteq \mathbb{R}^d$ be a compact topologically embedded C^2-hypersurface of $\mathbb{R}^d$ without boundary and let*

$$T(M,r) := \left\{x \in \mathbb{R}^d \;\middle|\; \inf_{y\in M} |x-y| \leqslant r\right\}, \quad r > 0,$$

be the tube of radius r around M. Then, there exists a constant $C_M > 0$ (that does not depend on r) such that for all $r \leqslant 1$ it holds that

$$\mathrm{vol}^d(T(M,r)) \leqslant C_M\, r. \tag{3.90}$$

Proof. The proof is based on Weyl's tube formula (Weyl, 1939). Let

$$\kappa := \max_{i\in\{1,\ldots,d-1\}} \kappa_i,$$

where κ_i is the i-th principal curvature of the hypersurface M (see (Gray, 2004, Section 3.1) for a formal definition). It follows from (Gray, 2004, Theorem 8.4 (i)) that

$$\mathrm{vol}^d(T(M,r)) = \sum_{i=0}^{\lfloor\frac{d-1}{2}\rfloor} \frac{2r^{2i+1}k_{2i}(M)}{\prod_{j=0}^{i}(1+2j)},$$

for all $r \leqslant \kappa^{-1}$, where $k_{2i}(M) = \int_M H_{2i}(x)\mathrm{d}x$, $i \in \{0,\ldots,\lfloor\frac{d-1}{2}\rfloor\}$, with H_{2i} denoting the so-called $(2i)$-th curvature of M, see (Gray, 2004, Section 4.1) for a formal definition. Now, thanks to M being a C^2-hypersurface, we have that H_{2i}, $i \in \{0,\ldots,\lfloor\frac{d-1}{2}\rfloor\}$, is bounded (see (Gray, 2004, Section 4.1)), which together with M compact (and thus bounded) implies $|k_{2i}(M)| < \infty$, for all $i \in \{0,\ldots,\lfloor\frac{d-1}{2}\rfloor\}$.

Moreover, by definition, $k_{2i}(M)$, $i \in \{0, \dots, \lfloor \frac{d-1}{2} \rfloor\}$, is independent of the tube radius r. Therefore, setting

$$C_M := \Big(\Big\lfloor \frac{d-1}{2} \Big\rfloor + 1 \Big) \max_i \frac{2|k_{2i}(M)|}{\prod_{j=0}^{i}(1+2j)}$$

establishes (3.90) for $0 < r \leqslant \min\{1, \kappa^{-1}\}$. It remains to prove (3.90) for $\min\{1, \kappa^{-1}\} < r \leq 1$. Let

$$R^* := \inf\{R > 0 \mid M \subseteq B_R(0)\}$$

and $D_{R^*} := \mathrm{vol}^d(B_{R^*+1}(0))$. Since

$$\mathrm{vol}^d(T(M,r)) \leq D_{R^*}, \quad \forall\, 0 < r \leq 1,$$

it follows that

$$\mathrm{vol}^d(T(M,r)) < D_{R^*} \max\{1, \kappa\}\, r,$$

for all $\min\{1, \kappa^{-1}\} < r \leq 1$, which establishes (3.90) for $\min\{1, \kappa^{-1}\} < r \leq 1$ and thereby concludes the proof. □

3.6.7. Proof of Proposition 6

We first upper-bound the integrand in $\|f - F_\tau f\|_2^2 = \int_{\mathbb{R}^d} |f(x) - f(x - \tau(x))|^2 \mathrm{d}x$. Owing to the mean value theorem (Comenetz, 2002, Theorem 3.7.5), we have

$$\begin{aligned} |f(x) - f(x - \tau(x))| &\leq \|\tau\|_\infty \sup_{y \in B_{\|\tau\|_\infty}(x)} |\nabla f(y)| \\ &\leqslant \underbrace{R\|\tau\|_\infty \sup_{y \in B_{\|\tau\|_\infty}(x)} \langle y \rangle^{-d}}_{=:h(x)}, \end{aligned}$$

where the last inequality follows by assumption. The idea is now to split the integral $\int_{\mathbb{R}^d} |h(x)|^2 \mathrm{d}x$ into integrals over the sets $B_1(0)$ and $\mathbb{R}^d \backslash B_1(0)$. For $x \in B_1(0)$, the monotonicity of the function $x \mapsto \langle x \rangle^{-d}$

implies $h(x) \leqslant R\|\tau\|_\infty \langle 0\rangle^{-d} = R\|\tau\|_\infty$, and for $x \in \mathbb{R}^d \backslash B_1(0)$, we have $(1 - \|\tau\|_\infty) \leqslant (1 - \frac{\|\tau\|_\infty}{|x|})$, which together with the monotonicity of $x \mapsto \langle x\rangle^{-d}$ yields $h(x) \leqslant R\|\tau\|_\infty \langle (1 - \frac{\|\tau\|_\infty}{|x|})x\rangle^{-d} \leqslant R\|\tau\|_\infty \langle (1 - \|\tau\|_\infty)x\rangle^{-d}$. Putting things together, we hence get

$$\begin{aligned} \|f - F_\tau f\|_2^2 &\leq R^2\|\tau\|_\infty^2 \Big(\mathrm{vol}^d\big(B_1(0)\big) + 2^d \int_{\mathbb{R}^d} \langle u\rangle^{-2d} \mathrm{d}u \Big) \qquad (3.91) \\ &\leq R^2\|\tau\|_\infty^2 \underbrace{\Big(\mathrm{vol}^d\big(B_1(0)\big) + 2^d \|\langle\cdot\rangle^{-d}\|_2^2 \Big)}_{=:D^2}, \end{aligned}$$

where in (3.91) we used the change of variables $u = (1 - \|\tau\|_\infty)x$, together with

$$\frac{\mathrm{d}u}{\mathrm{d}x} = (1 - \|\tau\|_\infty)^d \geqslant 2^{-d}. \qquad (3.92)$$

The inequality in (3.92) follows from $\|\tau\|_\infty < \frac{1}{2}$, which is by assumption. Since $\|\langle\cdot\rangle^{-d}\|_2 < \infty$, for $d \in \mathbb{N}$ (see, e.g., (Grafakos, 2008, Section 1)), and, obviously, $\mathrm{vol}^d\big(B_1(0)\big) < \infty$, it follows that $D^2 < \infty$. We finally get (3.29) with

$$C_R := RD, \qquad (3.93)$$

which completes the proof.

3.6.8. Proof of Proposition 7

Proposition 7. *Let $\Omega = \big((\Psi_n, M_n, P_n)\big)_{n\in\mathbb{N}}$ be an admissible module-sequence. The corresponding feature extractor $\Phi_\Omega : L^2(\mathbb{R}^d) \to (L^2(\mathbb{R}^d))^{\mathcal{Q}}$ is Lipschitz-continuous with Lipschitz constant $L_\Omega = 1$, i.e.,*

$$|||\Phi_\Omega(f) - \Phi_\Omega(h)||| \leqslant \|f - h\|_2, \quad \forall f, h \in L^2(\mathbb{R}^d). \qquad (3.94)$$

Remark 6. *Proposition 7 generalizes (Mallat, 2012, Proposition 2.5), which shows that the wavelet-modulus feature extractor Φ_W generated by scattering networks is Lipschitz-continuous with Lipschitz constant $L_W = 1$. Specifically, our generalization allows for*

general semi-discrete frames (i.e., general convolution filters), general Lipschitz-continuous non-linearities M_n, and general Lipschitz-continuous operators P_n, all of which can be different in different layers. Moreover, thanks to the admissibility condition (3.14)*, the Lipschitz constant $L_\Omega = 1$ in* (3.94) *is completely independent of the frame upper bounds B_n and the Lipschitz-constants L_n and R_n of M_n and P_n, respectively.*

Proof. The key idea of the proof is again—similarly to the proof of Proposition 3 in Section 3.6.1—to judiciously employ a telescoping series argument. For ease of notation, we let $f_q := U[q]f$ and $h_q := U[q]h$, for $f, h \in L^2(\mathbb{R}^d)$. Thanks to (3.11) and the admissibility condition (3.14), we have $\|f_q\|_2 \leqslant \|f\|_2 < \infty$ and $\|h_q\|_2 \leqslant \|h\|_2 < \infty$ and thus $f_q, h_q \in L^2(\mathbb{R}^d)$. We start by writing

$$
\begin{aligned}
|||\Phi_\Omega(f) - \Phi_\Omega(h)|||^2 &= \sum_{n=0}^{\infty} \sum_{q \in \Lambda^n} ||f_q * \chi_n - h_q * \chi_n||_2^2 \\
&= \lim_{N \to \infty} \sum_{n=0}^{N} \underbrace{\sum_{q \in \Lambda^n} ||f_q * \chi_n - h_q * \chi_n||_2^2}_{=:a_n} .
\end{aligned}
$$

As in the proof of Proposition 3 in Section 3.6.1, the key step is to show that a_n can be upper-bounded according to

$$
a_n \leqslant b_n - b_{n+1}, \qquad \forall\, n \in \mathbb{N}_0, \tag{3.95}
$$

where here $b_n := \sum_{q \in \Lambda^n} \|f_q - h_q\|_2^2$, for all $n \in \mathbb{N}_0$, and to note that, similarly to (3.35),

$$
\begin{aligned}
\sum_{n=0}^{N} a_n &\leqslant \sum_{n=0}^{N} (b_n - b_{n+1}) = (b_0 - b_1) + \cdots + (b_N - b_{N+1}) \\
&= b_0 - \underbrace{b_{N+1}}_{\geqslant 0} \leqslant b_0 = \sum_{q \in \Lambda^0} \|f_q - h_q\|_2^2 = \|U[e]f - U[e]h\|_2^2 \\
&= \|f - h\|_2^2,
\end{aligned}
$$

which then yields (3.94) according to

$$|||\Phi_\Omega(f) - \Phi_\Omega(h)|||^2 = \lim_{N\to\infty} \sum_{n=0}^{N} a_n \leqslant \lim_{N\to\infty} \|f-h\|_2^2 = \|f-h\|_2^2.$$

Writing out (3.95), it follows that we need to establish

$$\sum_{q\in\Lambda^n} \|f_q * \chi_n - h_q * \chi_n\|_2^2 \leqslant \sum_{q\in\Lambda^n} ||f_q - h_q||_2^2 - \sum_{q\in\Lambda^{n+1}} \|f_q - h_q\|_2^2, \quad (3.96)$$

for all $n \in \mathbb{N}_0$. We start by examining the second term on the RHS of (3.96) and note that, thanks to the decomposition

$$\tilde{q} \in \Lambda^{n+1} = \underbrace{\Lambda_1 \times \cdots \times \Lambda_n}_{=\Lambda^n} \times \Lambda_{n+1}$$

and $U[\tilde{q}] = U[(q, \lambda_{n+1})] = U_{n+1}[\lambda_{n+1}]U[q]$, by (3.10), we have

$$\begin{aligned} &\sum_{\tilde{q}\in\Lambda^{n+1}} \|f_{\tilde{q}} - h_{\tilde{q}}\|_2^2 \\ &= \sum_{q\in\Lambda^n} \sum_{\lambda_{n+1}\in\Lambda_{n+1}} \|U_{n+1}[\lambda_{n+1}]f_q - U_{n+1}[\lambda_{n+1}]h_q\|_2^2. \end{aligned} \quad (3.97)$$

Substituting (3.97) into (3.96) and rearranging terms, we obtain

$$\begin{aligned} &\sum_{q\in\Lambda^n} \Big(\|f_q * \chi_n - h_q * \chi_n\|_2^2 + \sum_{\lambda_{n+1}\in\Lambda_{n+1}} \|U_{n+1}[\lambda_{n+1}]f_q \\ &- U_{n+1}[\lambda_{n+1}]h_q\|_2^2 \Big) \leqslant \sum_{q\in\Lambda^n} ||f_q - h_q||_2^2, \end{aligned} \quad (3.98)$$

for all $n \in \mathbb{N}_0$. We next note that the second term inside the sum on the LHS of (3.98) satisfies

$$\begin{aligned} &\sum_{\lambda_{n+1}\in\Lambda_{n+1}} \|U_{n+1}[\lambda_{n+1}]f_q - U_{n+1}[\lambda_{n+1}]h_q\|_2^2 \\ &\leqslant \sum_{\lambda_{n+1}\in\Lambda_{n+1}} \|P_{n+1}\big(M_{n+1}(f_q * g_{\lambda_{n+1}})\big) \\ &- P_{n+1}\big(M_{n+1}(h_q * g_{\lambda_{n+1}})\big)\|_2^2, \end{aligned} \quad (3.99)$$

where we employed arguments similar to those leading to (3.39). Substituting the second term inside the sum on the LHS of (3.98) by the upper bound (3.99), and using the Lipschitz property of M_{n+1} and P_{n+1} yields

$$\sum_{q\in\Lambda^n}\Big(\|f_q*\chi_n-h_q*\chi_n\|_2^2+\sum_{\lambda_{n+1}\in\Lambda_{n+1}}\|U_{n+1}[\lambda_{n+1}]f_q - U_{n+1}[\lambda_{n+1}]h_q\|_2^2\Big)\leqslant\sum_{q\in\Lambda^n}\max\{1,L_{n+1}^2R_{n+1}^2\}\Big(\|(f_q-h_q)*\chi_n\|_2^2 + \sum_{\lambda_{n+1}\in\Lambda_{n+1}}\|(f_q-h_q)*g_{\lambda_{n+1}}\|_2^2\Big), \tag{3.100}$$

for all $n\in\mathbb{N}_0$. As the functions $\{g_{\lambda_{n+1}}\}_{\lambda_{n+1}\in\Lambda_{n+1}}\cup\{\chi_n\}$ are the atoms of the semi-discrete frame Ψ_{n+1} for $L^2(\mathbb{R}^d)$ and $f_q,h_q\in L^2(\mathbb{R}^d)$, as established above, we have

$$\|(f_q-h_q)*\chi_n\|_2^2+\sum_{\lambda_{n+1}\in\Lambda_{n+1}}\|(f_q-h_q)*g_{\lambda_{n+1}}\|_2^2\leqslant B_{n+1}\|f_q-h_q\|_2^2,$$

which, when used in (3.100) yields

$$\sum_{q\in\Lambda^n}\Big(\|f_q*\chi_n-h_q*\chi_n\|_2^2 + \sum_{\lambda_{n+1}\in\Lambda_{n+1}}\|U_{n+1}[\lambda_{n+1}]f_q-U_{n+1}[\lambda_{n+1}]h_q\|_2^2\Big)\leqslant\sum_{q\in\Lambda^n}\max\{B_{n+1},B_{n+1}L_{n+1}^2R_{n+1}^2\}\|f_q-h_q\|_2^2, \tag{3.101}$$

for all $n\in\mathbb{N}_0$. Finally, invoking the admissibility condition

$$\max\{B_n,B_nL_n^2R_n^2\}\leqslant 1,\qquad\forall n\in\mathbb{N},$$

in (3.101) we get (3.98) and hence (3.95). This completes the proof. □

CHAPTER 4

Energy propagation in deep convolutional neural networks

Many practical machine learning tasks employ very deep convolutional neural networks (He et al., 2015). Such large depths pose formidable computational challenges in training and operating the network. It is therefore important to understand how fast the energy contained in the propagated signals (a.k.a. feature maps) decays across layers. In addition, it is desirable that the feature extractor generated by the network be informative in the sense of the only signal mapping to the all-zeros feature vector being the zero input signal. This "trivial null-set" property can be accomplished by asking for "energy conservation" in the sense of the energy in the feature vector being proportional to that of the corresponding input signal. In this chapter, we establish conditions for energy conservation (and thus for a trivial null-set) for a wide class of DCNNs and characterize corresponding feature map energy decay rates. Specifically, we consider generalized scattering networks (introduced in Chapter 3) and find that under mild analyticity and high-pass conditions on the filters (which encompass, inter alia, various constructions of Weyl-Heisenberg filters, wavelets, ridgelets, (α)-curvelets, and shearlets) the feature map energy decays at least polynomially fast. For broad families of wavelets and Weyl-Heisenberg filters, the guaranteed decay rate is shown to be exponential. Moreover, we provide handy estimates of

the number of layers needed to have at least $((1-\varepsilon)\cdot 100)\%$ of the input signal energy be contained in the feature vector. Finally, we show how networks of fixed (possibly small) depth can be designed to capture most of the input signal's energy.

Outline

The remainder of this chapter is organized as follows. Section 4.1 presents the modulus-based scattering network architecture considered throughout this chapter. In Section 4.2, we formalize the notions of feature map energy decay and feature vector energy conservation, and present previous work on that topic. Section 4.3 contains our main results of this chapter, Theorems 3 and 4, which establish polynomial energy decay for general filters and exponential energy decay for structured filters (namely, for broad families of wavelets and Weyl-Heisenberg filters), respectively. Handy estimates of the number of layers needed to have most of the input signal energy be contained in the feature vector are provided in Section 4.4. Finally, in Section 4.5, we design scattering networks of fixed (possibly small) depth that capture most of the input signal's energy.

4.1. MODULUS-BASED NETWORKS

Throughout this chapter we consider (unless explicitly stated otherwise) input signals $f \in L^2(\mathbb{R}^d)$, and employ the module-sequence (see Definition 2 in Section 3.2)

$$\Omega := \big((\Psi_n, |\cdot|, \mathrm{Id})\big)_{n\in\mathbb{N}}, \tag{4.1}$$

i.e., each network layer is associated with (i) a collection of filters[1] $\Psi_n := \{\chi_{n-1}\} \cup \{g_{\lambda_n}\}_{\lambda_n\in\Lambda_n} \subseteq L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$, where χ_{n-1} and the

[1] We note that it is actually the notation $\Psi_n = \{T_b I\chi_{n-1}\}_{b\in\mathbb{R}^d} \cup \{T_b I g_{\lambda_n}\}_{b\in\mathbb{R}^d,\lambda_n\in\Lambda_n}$, rather than $\Psi_n = \{\chi_{n-1}\} \cup \{g_{\lambda_n}\}_{\lambda_n\in\Lambda_n}$ that was introduced in Definition 2 in Section 3.2, but in this chapter we prefer to work with $\Psi_n = \{\chi_{n-1}\} \cup \{g_{\lambda_n}\}_{\lambda_n\in\Lambda_n}$ for the sake of expositional simplicity.

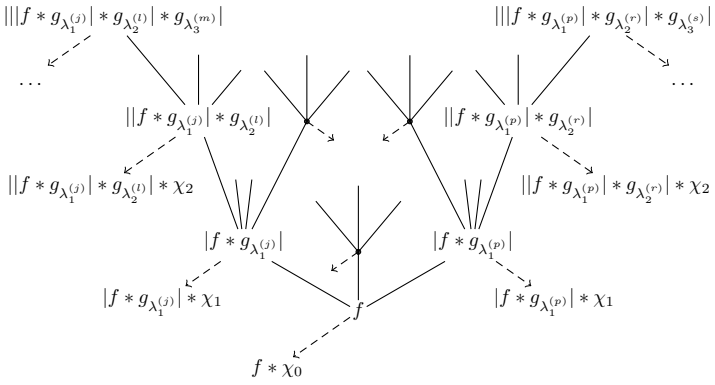

Fig. 4.1: Network architecture underlying the module sequence (4.1). The index $\lambda_n^{(k)}$ corresponds to the k-th filter $g_{\lambda_n^{(k)}}$ of the collection Ψ_n associated with the n-th network layer. The function χ_n is the output-generating filter of the n-th network layer. The root of the network corresponds to $n = 0$.

g_{λ_n}, indexed by a countable set Λ_n, satisfy the frame condition (2.1), i.e.,

$$A_n\|f\|_2^2 \leqslant \|f * \chi_{n-1}\|_2^2 + \sum_{\lambda_n \in \Lambda_n} \|f * g_{\lambda_n}\|^2 \leqslant B_n\|f\|_2^2, \qquad (4.2)$$

for all $f \in L^2(\mathbb{R}^d)$, for some $A_n, B_n > 0$, (ii) the modulus non-linearity $|\cdot| : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$, $|f|(x) = |f(x)|$ (see Section 2.3), and (iii) no pooling, which corresponds to pooling through the identity operator with pooling factors $S_n = 1$, for all $n \in \mathbb{N}$, see (3.5). Associated with the module $(\Psi_n, |\cdot|, \mathrm{Id})$, the operator $U_n[\lambda_n]$ defined in (3.7) particularizes to

$$U_n[\lambda_n]f = \big|f * g_{\lambda_n}\big|.$$

The feature maps $U[q]f$, $q \in \Lambda^n$, defined in (3.10), can therefore be written as

$$\begin{aligned} U[q]f &= U_n[\lambda_n] \cdots U_2[\lambda_2]U_1[\lambda_1]f \\ &= \big|\cdots\big|\,|f * g_{\lambda_1}| * g_{\lambda_2}\big| \cdots * g_{\lambda_n}\big|. \end{aligned} \qquad (4.3)$$

The architecture corresponding to the module sequence Ω in (4.1) is illustrated in Fig. 4.1

4.2. PROBLEM STATEMENT

The first central goal of this chapter is to understand how quickly the energy contained in the feature maps decays across layers. Specifically, we shall study the decay of

$$W_N(f) := \sum_{q \in \Lambda^N} \|U[q]f\|_2^2, \quad f \in L^2(\mathbb{R}^d), \tag{4.4}$$

as a function of network depth N. Moreover, it is desirable that the infinite-depth feature vector $\Phi_\Omega(f)$ be informative in the sense of the only signal mapping to the all-zeros feature vector being the zero input signal, i.e., Φ_Ω has a trivial null-set

$$\mathcal{N}(\Phi_\Omega) := \{f \in L^2(\mathbb{R}^d) \mid \Phi_\Omega(f) = 0\} \stackrel{!}{=} \{0\}. \tag{4.5}$$

Fig. 4.2 illustrates the practical ramifications of a non-trivial null-set in a binary classification task. $\mathcal{N}(\Phi_\Omega) = \{0\}$ can be guaranteed by asking for "energy conservation" in the sense of

$$A_\Omega \|f\|_2^2 \leqslant |||\Phi_\Omega(f)|||^2 \leqslant B_\Omega \|f\|_2^2, \quad \forall f \in L^2(\mathbb{R}^d), \tag{4.6}$$

for some constants $A_\Omega, B_\Omega > 0$ (possibly depending on the module-sequence Ω) and with the feature space norm $|||\Phi_\Omega(f)||| := \big(\sum_{n=0}^\infty |||\Phi_\Omega^n(f)|||^2\big)^{1/2}$, where $|||\Phi_\Omega^n(f)||| := \big(\sum_{q \in \Lambda^n} \|(U[q]f) * \chi_n\|_2^2\big)^{1/2}$. Indeed, (4.5) follows from (4.6) as the upper bound in (4.6) yields $\{0\} \subseteq \mathcal{N}(\Phi_\Omega)$, and the lower bound implies $\{0\} \supseteq \mathcal{N}(\Phi_\Omega)$. We emphasize that, as Φ_Ω is a non-linear operator (owing to the modulus non-linearities), characterizing its null-set is non-trivial in general. The upper bound in (4.6) was established in Section 3.6.1. While the existence of this upper bound is implied by the filters Ψ_n, $n \in \mathbb{N}$, satisfying the frame property (4.2), perhaps surprisingly, this is not enough to guarantee $A_\Omega > 0$ (see Section 4.6 for an illustrative

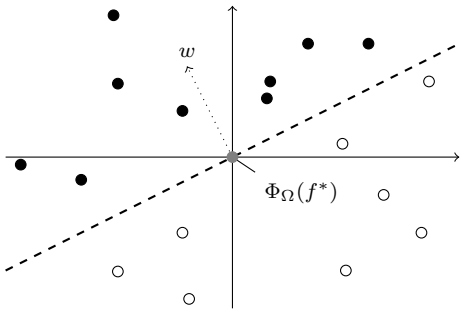

Fig. 4.2: Impact of a non-trivial null-set $\mathcal{N}(\Phi_\Omega)$ in a binary classification task. The feature vector $\Phi_\Omega(f)$ is fed into a linear classifier Bishop (2009), which determines set membership based on the sign of the inner product $\langle w, \Phi_\Omega(f)\rangle$. The (learned) weight vector w is perpendicular to the separating hyperplane (dashed line). If the null-set of the feature extractor Φ_Ω is non-trivial, there exist input signals $f^* \neq 0$ that are mapped to the origin in feature space, i.e., $\Phi_\Omega(f^*) = 0$ (gray circle), and therefore lie—independently of the weight vector w—on the separating hyperplane. These input signals $f^* \neq 0$ are therefore unclassifiable.

example). We refer the reader to Section 4.4 for results on the null-set of the *finite-depth* feature extractor $\bigcup_{n=0}^{N} \Phi_\Omega^n$. Finally, we emphasize that throughout the thesis energy decay results pertain to the feature maps $U[q]f$, whereas energy conservation according to (4.6) applies to the feature vector $\Phi_\Omega(f)$.

Previous work on the decay rate of $W_N(f)$ in (Waldspurger, 2015, Section 5) shows that for wavelet-based networks (i.e., in every network layer, the filters $\Psi = \{\chi\} \cup \{g_\lambda\}_{\lambda\in\Lambda}$ in (4.1) are taken to be (specific) 1-D wavelets that constitute a Parseval frame, with χ a low-pass filter) there exist $\varepsilon > 0$ and $a > 1$ (both constants unspecified) such that

$$W_N(f) \leqslant \int_{\mathbb{R}} |\widehat{f}(\omega)|^2 \Big(1 - \Big|\widehat{r_g}\Big(\frac{\omega}{\varepsilon a^{N-1}}\Big)\Big|^2\Big) \mathrm{d}\omega, \tag{4.7}$$

for real-valued 1-D signals $f \in L^2(\mathbb{R})$ and $N \geqslant 2$, where $\widehat{r_g}(\omega) := e^{-\omega^2}$. To see that this result indicates energy decay, Fig. 4.3 illustrates the

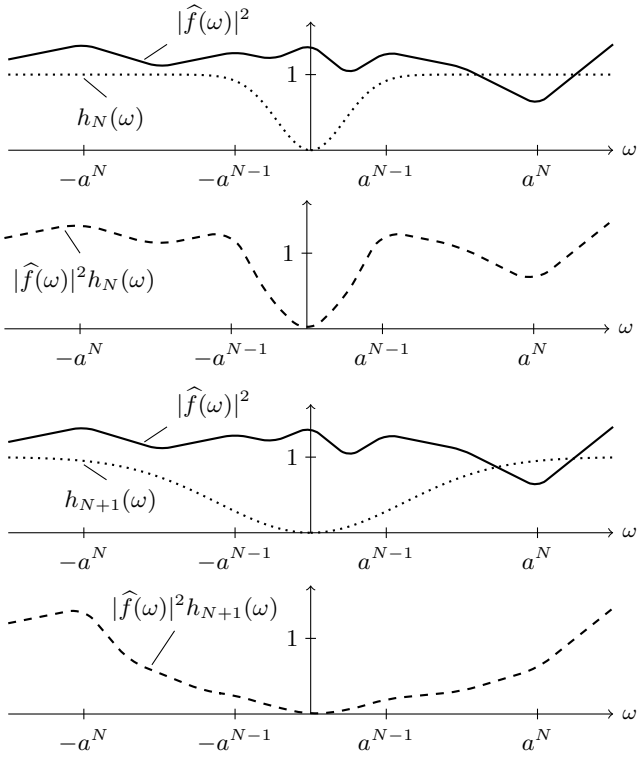

Fig. 4.3: Illustration of the impact of network depth N on the upper bound on $W_N(f)$ in (4.7), for $\varepsilon = 1$. The function $h_N(\omega) := (1 - \widehat{r_g}(\frac{\omega}{\varepsilon a^{N-1}}))$, where $\widehat{r_g}(\omega) = e^{-\omega^2}$, is of increasing high-pass nature as N increases, which results in cutting out increasing amounts of low-frequency energy of f and thereby making the upper bound in (4.7) decay as a function of N.

influence of network depth N on the upper bound in (4.7). Specifically, we can see that increasing the network depth results in cutting out increasing amounts of low-frequency energy of f and thereby making the upper bound in (4.7) decay as a function of N. Moreover, it is interesting to note that the upper bound on $W_N(f) = \sum_{q \in \Lambda^N} \|U[q]f\|_2^2$

is independent of the wavelets generating the feature maps $U[q]f$, $q \in \Lambda^N$. For scattering networks that employ, in every network layer, uniform covering filters $\Psi = \{\chi\} \cup \{g_\lambda\}_{\lambda\in\Lambda} \subseteq L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$ forming a Parseval frame (where χ, again, is a low-pass filter), exponential energy decay according to

$$W_N(f) = \mathcal{O}(a^{-N}), \quad \forall f \in L^2(\mathbb{R}^d), \tag{4.8}$$

for an unspecified $a > 1$, was established in (Czaja and Li, 2017, Proposition 3.3). Moreover, (Waldspurger, 2015, Section 5) and (Czaja and Li, 2017, Theorem 3.6 (a)) state—for the respective module-sequences—that (4.6) holds with $A_\Omega = B_\Omega = 1$ and hence

$$|||\Phi_\Omega(f)|||^2 = \|f\|_2^2. \tag{4.9}$$

The first main goal of this chapter is to establish i) for d-dimensional complex-valued input signals that (4.4) decays polynomially according to

$$W_N(f) \leqslant B_\Omega^N \int_{\mathbb{R}^d} \big|\widehat{f}(\omega)\big|^2 \Big(1 - \Big|\widehat{r}_l\Big(\frac{\omega}{N^\alpha}\Big)\Big|^2\Big)\mathrm{d}\omega, \tag{4.10}$$

for all $f \in L^2(\mathbb{R}^d)$ and all $N \geqslant 1$, where

$$\alpha = \begin{cases} 1, & d = 1, \\ \log_2(\sqrt{d/(d-1/2)}), & d \geqslant 2, \end{cases}$$

$B_\Omega^N = \prod_{k=1}^N \max\{1, B_k\}$, and $\widehat{r}_l : \mathbb{R}^d \to \mathbb{R}$, $\widehat{r}_l(\omega) = (1 - |\omega|)_+^l$, with $l > \lfloor d/2 \rfloor + 1$, for networks based on general filters $\{\chi_{n-1}\} \cup \{g_{\lambda_n}\}_{\lambda_n\in\Lambda_n}$ that satisfy mild analyticity and high-pass conditions and are allowed to be different in different network layers (with the proviso that χ_{n-1}, $n \in \mathbb{N}$, is of low-pass nature in a sense to be made precise), and ii) for 1-D complex-valued input signals that (4.4) decays exponentially according to

$$W_N(f) \leqslant \int_{\mathbb{R}} \big|\widehat{f}(\omega)\big|^2 \Big(1 - \Big|\widehat{r}_l\Big(\frac{\omega}{a^{N-1}}\Big)\Big|^2\Big)\mathrm{d}\omega, \tag{4.11}$$

for all $f \in L^2(\mathbb{R}^d)$ and all $N \geqslant 1$, for networks that are based, in every network layer, on a broad family of wavelets or on a broad family

of Weyl-Heisenberg filters. Here, we emphasize that an arbitrary decay factor $a > 1$ can be realized through suitable choice of the mother wavelet bandwidth or the Weyl-Heisenberg prototype function bandwidth. Thanks to the RHS of (4.10) and (4.11) not depending on the specific filters $\{\chi_{n-1}\} \cup \{g_{\lambda_n}\}_{\lambda_n \in \Lambda_n}$, we will be able to establish—under smoothness assumptions on the input signal f—universal energy decay results. Specifically, particularizing the RHS in (4.10) and (4.11) to Sobolev-class input signals $f \in H^s(\mathbb{R}^d)$, $s > 0$, where

$$H^s(\mathbb{R}^d) = \Big\{ f \in L^2(\mathbb{R}^d) \;\Big|\; \int_{\mathbb{R}^d} (1 + |\omega|^2)^s |\widehat{f}(\omega)|^2 \mathrm{d}\omega < \infty \Big\},$$

we show that (4.10) yields polynomial energy decay according to

$$W_N(f) = \mathcal{O}\big(N^{-\gamma\alpha}\big), \quad \forall f \in H^s(\mathbb{R}^d), \tag{4.12}$$

and (4.11) exponential energy decay according to

$$\mathcal{O}\big(a^{-\gamma N}\big), \quad \forall f \in H^s(\mathbb{R}), \tag{4.13}$$

where $\gamma := \min\{1, 2s\}$ in both cases. Sobolev spaces $H^s(\mathbb{R}^d)$ contain a wide range of practically relevant signal classes such as, e.g.,

i) the space $L^2_L(\mathbb{R}^d) = \{f \in L^2(\mathbb{R}^d) \mid \mathrm{supp}(\widehat{f}\,) \subseteq B_L(0)\}$, $L \geqslant 0$, of L-band-limited signals according to $L^2_L(\mathbb{R}^d) \subseteq H^s(\mathbb{R}^d)$, for all $L \geqslant 0$ and all $s > 0$, which follows from

$$\begin{aligned} \int_{\mathbb{R}^d} (1 + |\omega|^2)^s |\widehat{f}(\omega)|^2 \mathrm{d}\omega &= \int_{B_L(0)} (1 + |\omega|^2)^s |\widehat{f}(\omega)|^2 \mathrm{d}\omega \\ &\leqslant (1 + |L|^2)^s \|f\|_2^2 < \infty, \end{aligned}$$

for $f \in L^2_L(\mathbb{R}^d)$, $L \geqslant 0$, and $s > 0$, where we used Parseval's formula and the fact that $\omega \mapsto (1+|\omega|^2)^s$, $\omega \in \mathbb{R}^d$, is monotonically increasing in $|\omega|$, for all $s > 0$,

ii) the space $\mathcal{C}^K_{\mathrm{CART}}$ of cartoon functions of size K, introduced in (Donoho, 2001), and widely used in the mathematical signal processing literature (Kutyniok and Labate, 2012a; Grohs and Kutyniok,

2014; Grohs et al., 2015) as a model for natural images such as, e.g., images of handwritten digits (LeCun and Cortes, 1998) (see Fig. 3.7). For a formal definition of $\mathcal{C}^K_{\mathrm{CART}}$, we refer the reader to Section 3.4.3. In Section 4.7.1 we show that $\mathcal{C}^K_{\mathrm{CART}} \subseteq H^s(\mathbb{R}^d)$, for all $K > 0$ and all $s \in (0, 1/2)$.

Moreover, Sobolev functions are contained in the space of k-times continuously differentiable functions $C^k(\mathbb{R}^d, \mathbb{C})$ according to $H^s(\mathbb{R}^d) \subseteq C^k(\mathbb{R}^d, \mathbb{C})$, for all $s > k + \frac{d}{2}$ (Adams, 1975, Section 4).

Our second central goal in this chapter is to establish energy conservation according to (4.6) (which, as explained above, implies $\mathcal{N}(\Phi_\Omega) = \{0\}$) for the network configurations corresponding to the energy decay results (4.10) and (4.11). Finally, we provide handy estimates of the number of layers needed to have at least $((1-\varepsilon)\cdot 100)\%$ of the input signal energy be contained in the feature vector.

4.3. ENERGY DECAY AND CONSERVATION

Throughout Chapter 4, we make the following assumptions on the filters $\{g_{\lambda_n}\}_{\lambda_n \in \Lambda_n}$.

Assumption 1. *The $\{g_{\lambda_n}\}_{\lambda_n \in \Lambda_n}$, $n \in \mathbb{N}$, are analytic in the following sense: For every layer index $n \in \mathbb{N}$, for every $\lambda_n \in \Lambda_n$, there exists an orthant $H_{A_{\lambda_n}} \subseteq \mathbb{R}^d$, with $A_{\lambda_n} \in O(d)$, such that*

$$\mathrm{supp}(\widehat{g_{\lambda_n}}) \subseteq H_{A_{\lambda_n}}. \tag{4.14}$$

Moreover, there exists $\delta > 0$ such that

$$\sum_{\lambda_n \in \Lambda_n} |\widehat{g_{\lambda_n}}(\omega)|^2 = 0, \quad \text{a.e. } \omega \in B_\delta(0). \tag{4.15}$$

In the 1-D case, i.e., for $d = 1$, Assumption 1 simply amounts to every filter g_{λ_n} satisfying

$$\text{either} \quad \mathrm{supp}(\widehat{g_{\lambda_n}}) \subseteq (-\infty, -\delta] \quad \text{or} \quad \mathrm{supp}(\widehat{g_{\lambda_n}}) \subseteq [\delta, \infty),$$

which constitutes an "analyticity" and "high-pass" condition. For dimensions $d \geqslant 2$, Assumption 1 requires that every filter g_{λ_n} be of high-pass nature and have a Fourier transform supported in a (not necessarily canonical) orthant. Since the frame condition (4.2) is equivalent to the Littlewood-Paley condition (2.3) (see Proposition 1), i.e.,

$$A_n \leqslant |\widehat{\chi_{n-1}}(\omega)|^2 + \sum_{\lambda_n \in \Lambda_n} |\widehat{g_{\lambda_n}}(\omega)|^2 \leqslant B_n, \quad \text{a.e. } \omega \in \mathbb{R}^d, \tag{4.16}$$

(4.15) implies low-pass characteristics for χ_{n-1} to fill the spectral gap $B_\delta(0)$ left by the filters $\{g_{\lambda_n}\}_{\lambda_n \in \Lambda_n}$.

The conditions (4.14) and (4.15) we impose on the Ψ_n, $n \in \mathbb{N}$, are not overly restrictive as they encompass, inter alia, various constructions of Weyl-Heisenberg filters (e.g., with a prototype function whose Fourier transform is a 1-D B-spline (Gröchenig et al., 2003, Section 1)), wavelets (e.g., analytic Meyer wavelets (Daubechies, 1992, Section 3.3.5) in 1-D, and Cauchy wavelets (Vandergheynst, 2002b) in 2-D), and specific constructions of ridgelets (Grohs, 2012, Section 2.2), curvelets (Candès and Donoho, 2005, Section 4.1), α-curvelets (Grohs et al., 2015, Section 3), and shearlets (e.g., cone-adapted shearlets (Kutyniok and Labate, 2012a, Section 4.3)). We refer the reader to Sections 2.2.1 and 2.2.2 for a brief review of some of these filter structures.

4.3.1. Polynomial energy decay

We are now ready to state our first main result on energy decay and energy conservation.

Theorem 3. *Let Ω be the module-sequence (4.1) with filters $\{g_{\lambda_n}\}_{\lambda_n \in \Lambda_n}$ satisfying the conditions in Assumption 1, and let $\delta > 0$ be the radius of the spectral gap $B_\delta(0)$ left by the filters $\{g_{\lambda_n}\}_{\lambda_n \in \Lambda_n}$ according to (4.15). Furthermore, let $A_\Omega^N := \prod_{k=1}^N \min\{1, A_k\}$, $B_\Omega^N := \prod_{k=1}^N \max\{1, B_k\}$, and*

$$\alpha := \begin{cases} 1, & d = 1, \\ \log_2(\sqrt{d/(d-1/2)}), & d \geqslant 2. \end{cases} \tag{4.17}$$

i) We have

$$W_N(f) \leqslant B_\Omega^N \int_{\mathbb{R}^d} \left|\widehat{f}(\omega)\right|^2 \Big(1 - \Big|\widehat{r}_l\Big(\frac{\omega}{N^\alpha \delta}\Big)\Big|^2\Big) \mathrm{d}\omega, \tag{4.18}$$

for all $f \in L^2(\mathbb{R}^d)$ and all $N \geqslant 1$, where $\widehat{r}_l : \mathbb{R}^d \to \mathbb{R}$, $\widehat{r}_l(\omega) := (1 - |\omega|)_+^l$, with $l > \lfloor d/2 \rfloor + 1$.

ii) For every Sobolev function $f \in H^s(\mathbb{R}^d)$, $s > 0$, we have

$$W_N(f) = \mathcal{O}\big(B_\Omega^N N^{-\gamma\alpha}\big), \tag{4.19}$$

where $\gamma := \min\{1, 2s\}$.

iii) If, in addition to Assumption 1,

$$0 < A_\Omega := \lim_{N\to\infty} A_\Omega^N \leqslant B_\Omega := \lim_{N\to\infty} B_\Omega^N < \infty, \tag{4.20}$$

then we have energy conservation according to

$$A_\Omega \|f\|_2^2 \leqslant |||\Phi_\Omega(f)|||^2 \leqslant B_\Omega \|f\|_2^2, \qquad \forall\, f \in L^2(\mathbb{R}^d). \tag{4.21}$$

Proof. For the proofs of i) and ii), we refer to the Sections 4.7.2 and 4.7.3, respectively. The proof of statement iii) is based on two key ingredients. First, we establish—in Proposition 8 in Section 4.7.4—that the feature extractor Φ_Ω satisfes the energy decomposition identity

$$A_\Omega^N \|f\|_2^2 \leqslant \sum_{n=0}^{N-1} |||\Phi_\Omega^n(f)|||^2 + W_N(f) \leqslant B_\Omega^N \|f\|_2^2, \tag{4.22}$$

for all $f \in L^2(\mathbb{R}^d)$ and all $N \geqslant 1$. Second, we show—in Proposition 9 in Section 4.7.5—that the integral on the RHS of (4.18) goes to zero as $N \to \infty$ which, thanks to $\lim_{N\to\infty} B_\Omega^N = B_\Omega < \infty$, implies that $W_N(f) \to 0$ as $N \to \infty$. We note that while the decomposition (4.22) holds for general filters $\{g_{\lambda_n}\}_{\lambda_n \in \Lambda_n}$ satisfying the frame property (4.2), it is the upper bound (4.18) that makes use of the analyticity and high-pass conditions in Assumption 1. The final energy conservation result (4.21) is obtained by letting $N \to \infty$ in (4.22). □

The strength of the results in Theorem 3 derives itself from the fact that the only condition we need to impose on the filters $\{g_{\lambda_n}\}_{\lambda_n \in \Lambda_n}$ is Assumption 1, which as already mentioned, is met by a wide array of filters. Moreover, condition (4.20) is easily satisfied by normalizing the filters Ψ_n, $n \in \mathbb{N}$, appropriately (see, e.g., Proposition 2 in Section 2.2). We note that this normalization, when applied to filters that satisfy Assumption 1, yields filters that still meet Assumption 1.

The identity (4.19) establishes, upon normalization (see, e.g., Proposition 2 in Section 2.2) of the Ψ_n to get $B_n \leqslant 1$, $n \in \mathbb{N}$, that the energy decay rate, i.e., the decay rate of $W_N(f)$, is at least polynomial in N. We hasten to add that (4.19) does not preclude the energy from decaying faster in practice.

Underlying the energy conservation result (4.21) is the following demodulation effect induced by the modulus non-linearity in combination with the analyticity and high-pass properties of the filters $\{g_{\lambda_n}\}_{\lambda_n \in \Lambda_n}$. In every network layer, the spectral content of each individual feature map is moved to base-band (i.e., to low frequencies), where it is extracted by the low-pass output-generating atom χ_n, see Fig. 4.4. The components not collected by χ_n (see Fig. 4.4, bottom row) are captured by the analytic high-pass filters $\{g_{\lambda_{n+1}}\}_{\lambda_{n+1} \in \Lambda_{n+1}}$ in the next layer and, thanks to the modulus non-linearity, again moved to low frequencies and extracted by χ_{n+1}. Iterating this process ensures that the null-set of the feature vector (be it for the infinite-depth network or, as established in Section 4.4, for finite network depths) is trivial. It is interesting to observe that the sigmoid, the rectified linear unit, and the hyperbolic tangent non-linearities—all widely used in the deep learning literature—exhibit very different behavior in this regard, namely, they do not demodulate in the way the modulus non-linearity does (Wiatowski et al., 2017, Figure 6). It is therefore unclear whether the proof machinery for energy conservation developed in this thesis extends to these non-linearities or, for that matter, whether one gets energy decay and conservation at all.

The feature map energy decay result (4.19) relates to the feature vector energy conservation result (4.21) via the energy decomposition identity (4.22). Specifically, particularizing (4.22) for Parseval frames,

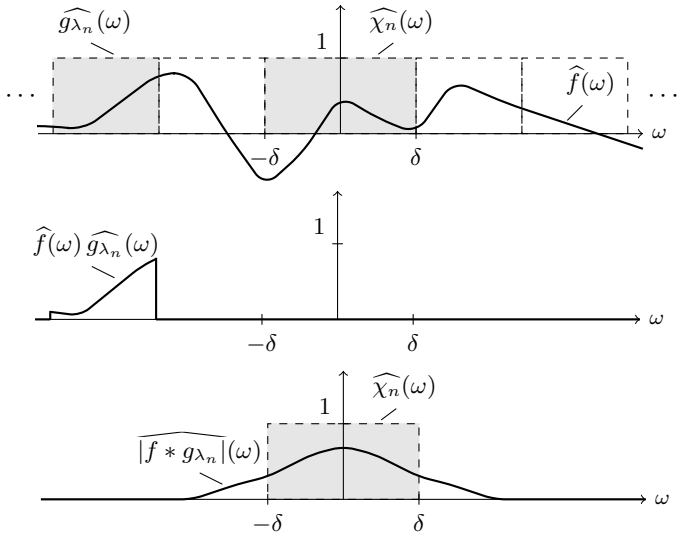

Fig. 4.4: Illustration of the demodulation effect of the modulus non-linearity. The $\{g_{\lambda_n}\}_{\lambda_n \in \Lambda_n}$ are taken as perfect band-pass filters (e.g., band-limited analytic Weyl-Heisenberg filters) and hence trivially satisfy the conditions in Assumption 1. The modulus operation in combination with the analyticity and the high-pass nature of the filters $\{g_{\lambda_n}\}_{\lambda_n \in \Lambda_n}$ ensures that—in every network layer—the spectral content of each individual feature map is moved to base-band (i.e., to low frequencies), where it is extracted by the (low-pass) output-generating filter χ_n.

i.e., $A_n = B_n = 1$, for all $n \in \mathbb{N}$, we get

$$\sum_{n=0}^{N-1} |||\Phi_\Omega^n(f)|||^2 + W_N(f) = \|f\|_2^2. \tag{4.23}$$

This shows that the input signal energy contained in the network layers $n \geqslant N$ is precisely given by $W_N(f)$. Thanks to $W_N(f) \to 0$ as $N \to \infty$ (established in Proposition 9 in Section 4.7.5) this residual energy will eventually be collected in the infinite-depth feature vector $\Phi_\Omega(f)$ so that no input signal energy is "lost" in the network. In Section 4.4, we shall answer the question of how many layers are

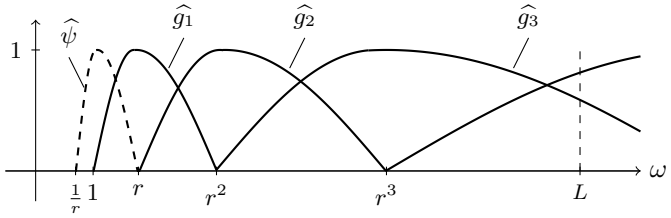

Fig. 4.5: Illustration of the Fourier transforms of the wavelet filters g_j on the frequency band $[0, L]$. The Fourier transform $\widehat{\psi}$ of the mother wavelet ψ is supported on the interval $[r^{-1}, r]$.

needed to absorb $((1-\varepsilon)\cdot 100)\%$ of the input signal energy.

4.3.2. Exponential energy decay

The next result shows that, under additional structural assumptions on the filters $\{g_{\lambda_n}\}_{\lambda_n \in \Lambda}$, the guaranteed energy decay rate can be improved from polynomial to exponential. Specifically, we construct 1-D wavelets and 1-D Weyl-Heisenberg filters that realize exponential energy decay according to $W_n(f) = \mathcal{O}(a^{-n})$, with arbitrary $a > 1$. Moreover, we want to tune the decay factor a by adjusting a single parameter, which will be seen to determine the mother wavelet or the Weyl-Heisenberg prototype function bandwidth. This will be accomplished through the following constructions:

i) *Wavelets*: For fixed $r > 1$, let the mother and father wavelets $\psi, \phi \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ satisfy the Littlewood-Paley condition

$$|\widehat{\phi}(\omega)|^2 + \sum_{j=1}^{\infty} |\widehat{\psi}(r^{-j}\omega)|^2 = 1, \qquad a.e.\ \omega \geqslant 0, \tag{4.24}$$

with $\mathrm{supp}(\widehat{\psi}) = [r^{-1}, r]$ and $\widehat{\psi}$ real-valued. Moreover, let $g_j(x) := r^j \psi(r^j x)$, $j \geqslant 1$, $g_j(x) := r^{|j|}\psi(-r^{|j|}x)$, $j \leqslant -1$, and let the output-generating filter be $\chi(x) := \phi(|x|)$, $x \in \mathbb{R}$. The Fourier transforms of the wavelets g_j and the mother wavelet ψ are illustrated in Fig. 4.5.

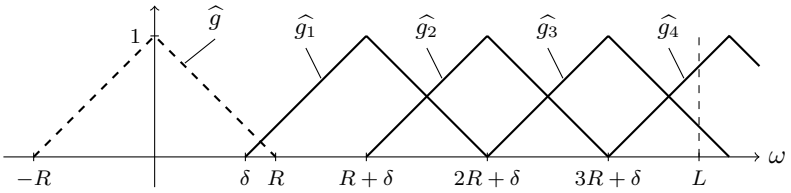

Fig. 4.6: Illustration of the Fourier transforms of the Weyl-Heisenberg filters g_k on the frequency band $[0, L]$. The Fourier transform $\widehat{g}$ of the prototype function g is supported on the interval $[-R, R]$.

ii) *Weyl-Heisenberg filters*: For fixed $R > 0$, $\delta \geqslant \frac{R}{2}$, let the functions $g, \phi \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ satisfy the Littlewood-Paley condition

$$|\widehat{\phi}(\omega)|^2 + \sum_{k=1}^{\infty} |\widehat{g}(\omega - (Rk + \delta))|^2 = 1, \qquad a.e.\ \omega \geqslant 0, \tag{4.25}$$

with $\operatorname{supp}(\widehat{g}) = [-R, R]$, $\widehat{g}(-\omega) = \widehat{g}(\omega)$, and $\widehat{g}$ real-valued. Moreover, let $g_k(x) := e^{2\pi i(Rk+\delta)x} g(x)$, $k \geqslant 1$, $g_k(x) := e^{-2\pi i(R|k|+\delta)x} g(x)$, $k \leqslant -1$, and set $\chi(x) := \phi(|x|)$, $x \in \mathbb{R}$. The Fourier transforms $\widehat{g_k}$ and $\widehat{g}$ are illustrated in Fig. 4.6.

The conditions we impose can be satisfied by constructing ψ, ϕ in i) from, e.g., an analytic Meyer wavelet (Daubechies, 1992, Section 3.3.5), and g, ϕ in ii) from a function whose Fourier transform is a 1-D B-spline (Gröchenig et al., 2003, Section 1). We emphasize that both the wavelet and Weyl-Heisenberg filters satisfy—by construction—the analyticity and highpass condition in Assumption 1.

We next state our main result on exponential feature map energy decay. For simplicity of exposition, we employ filters that are identical across network layers.

Theorem 4. *Let* $\widehat{r}_l : \mathbb{R} \to \mathbb{R}$, $\widehat{r}_l(\omega) := (1 - |\omega|)_+^l$, *with* $l > 1$.

i) Wavelets: Let $r > 1$, *set*

$$a := \frac{r^2 + 1}{r^2 - 1}, \tag{4.26}$$

and let Ω *be the module-sequence* (4.1) *with filters* $\Psi = \{\chi\} \cup \{g_j\}_{j\in\mathbb{Z}\setminus\{0\}}$ *in every network layer. Then,*

$$W_N(f) \leqslant \int_{\mathbb{R}} \left|\widehat{f}(\omega)\right|^2 \Big(1 - \Big|\widehat{r_l}\Big(\frac{\omega}{a^{N-1}}\Big)\Big|^2\Big)\mathrm{d}\omega, \tag{4.27}$$

for all $f \in L^2(\mathbb{R})$ *and all* $N \geqslant 1$*. Moreover, for every Sobolev function* $f \in H^s(\mathbb{R})$*,* $s > 0$*, we have*

$$W_N(f) = \mathcal{O}\big(a^{-\gamma N}\big), \tag{4.28}$$

where $\gamma := \min\{1, 2s\}$*.*

ii) Weyl-Heisenberg filters: Let $R > 0$*,* $\delta \geqslant \frac{R}{2}$*, set*

$$a := \frac{1}{2} + \frac{\delta}{R}, \tag{4.29}$$

and let Ω *be the module-sequence* (4.1) *with filters* $\Psi = \{\chi\} \cup \{g_k\}_{k\in\mathbb{Z}\setminus\{0\}}$ *in every network layer. Then,*

$$W_N(f) \leqslant \int_{\mathbb{R}} \left|\widehat{f}(\omega)\right|^2 \Big(1 - \Big|\widehat{r_l}\Big(\frac{\omega}{a^{N-1}\delta}\Big)\Big|^2\Big)\mathrm{d}\omega, \tag{4.30}$$

for all $f \in L^2(\mathbb{R})$ *and all* $N \geqslant 1$*. Moreover, for every Sobolev function* $f \in H^s(\mathbb{R})$*,* $s > 0$*, we have*

$$W_N(f) = \mathcal{O}\big(a^{-\gamma N}\big), \tag{4.31}$$

where $\gamma := \min\{1, 2s\}$*.*

Proof. The proof is given in Section 4.7.6. □

The identities (4.26) and (4.29) show that the filter constructions we propose, indeed, allow to tune the decay factor a through a single parameter, namely r in the wavelet case and R in the Weyl-Heisenberg case. Reducing r, R results in faster energy decay (see also Fig. 4.7). In addition, we note that in the presence of pooling by sub-sampling, say with pooling factors $S_n := S \in [1, a)$, for all $n \in \mathbb{N}$, the effective decay factor in (4.28) and (4.31) becomes $\frac{a}{S}$. Hence, exponential energy

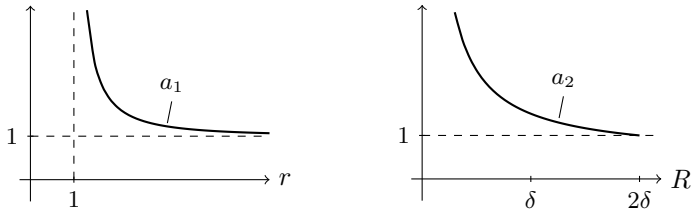

Fig. 4.7: Illustration of the functions $a_1(r) := \frac{r^2+1}{r^2-1}$, for $r > 1$, (left plot) and $a_2(R) := \frac{1}{2} + \frac{\delta}{R}$, for $R \leqslant 2\delta$, (right plot).

decay is compatible with vertical translation invariance according to Theorem 1 in Section 3.3, albeit at the cost of slower (exponential) decay. The proof of this statement is structurally very similar to that of Theorem 4 and will therefore not be presented here. We next put the results in Theorems 3 and 4 into perspective w.r.t. to the literature.

4.3.3. Relation to the literature

Relation to (Waldspurger, 2015, Section 5)

The basic philosophy of our proof technique for (4.18), (4.21), (4.27), and (4.30) is inspired by the proof in (Waldspurger, 2015, Section 5), which establishes (4.7) and (4.9) for scattering networks based on certain wavelet filters and with 1-D real-valued input signals $f \in L^2(\mathbb{R})$. Specifically, in (Waldspurger, 2015, Section 5), in every network layer, the filters $\Psi_W = \{\chi\} \cup \{g_j\}_{j\in\mathbb{Z}}$ (where $g_j(\omega) := 2^j\psi(2^j\omega)$, $j \in \mathbb{Z}$, for some mother wavelet $\psi \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$) are 1-D functions satisfying the frame property (4.2) with $A_n = B_n = 1$, $n \in \mathbb{N}$, a mild analyticity condition (Waldspurger, 2015, Equation 5.5) in the sense of $|\widehat{g_j}(\omega)|$, $j \in \mathbb{Z}$, being larger for positive frequencies ω than for the corresponding negative ones, and a vanishing moments condition (Waldspurger, 2015, Equation 5.6) which controls the behavior of $\widehat{\psi}(\omega)$ around the origin according to $|\widehat{\psi}(\omega)| \leqslant C|\omega|^{1+\varepsilon}$, $\omega \in \mathbb{R}$, for

some $C, \varepsilon > 0$. Similarly to the proof of (4.9) in (Waldspurger, 2015, Section 5), we base our proof of (4.21) on the energy decomposition identity (4.22) and on an upper bound on $W_N(f)$ (see (4.7) for the corresponding upper bound established in (Waldspurger, 2015, Section 5)) shown to go to zero as $N \to \infty$. The explicit energy decay results (4.19), (4.28), and (4.31) for $f \in H^s(\mathbb{R}^d)$ are entirely new. The major differences between (Waldspurger, 2015, Section 5) and our results are (i) that (4.7) (reported in (Waldspurger, 2015, Section 5)) depends on an unspecified $a > 1$, whereas our results in (4.18), (4.19), (4.27), (4.28), (4.30), and (4.31) make the decay factor a and the decay exponent α explicit, (ii) the technical elements employed to arrive at the upper bounds on $W_N(f)$, specifically, while the proof in (Waldspurger, 2015, Section 5) makes explicit use of the algebraic structure of the filters, namely, the multi-scale structure of wavelets, our proof of (4.18) is oblivious to the algebraic structure of the filters, which is why it applies to general (possibly unstructured) filters that, in addition, can be different in different network layers, (iii) the assumptions imposed on the filters, namely the analyticity and vanishing moments conditions in (Waldspurger, 2015, Equations 5.5–5.6), in contrast to our Assumption 1, and (iv) the class of input signals f the results apply to, namely 1-D real-valued signals in (Waldspurger, 2015, Section 5), and d-dimensional complex-valued signals in our Theorem 3 in Section 4.3.1.

Relation to (Czaja and Li, 2017)

For scattering networks that are based on so-called uniform covering filters (Czaja and Li, 2017), (4.8) and (4.9) are established in (Czaja and Li, 2017) for d-dimensional complex-valued input signals $f \in L^2(\mathbb{R}^d)$. Specifically, in (Czaja and Li, 2017), in every network layer, the d-dimensional filters $\{\chi\} \cup \{g_\lambda\}_{\lambda \in \Lambda}$ are taken to satisfy i) the frame property (4.2) with $A = B = 1$ and hence $A_n = B_n = 1$, $n \in \mathbb{N}$, see (Czaja and Li, 2017, Definition 2.1 (c)), ii) a vanishing moments condition (Czaja and Li, 2017, Definition 2.1 (a)) according to $\widehat{g_\lambda}(0) = 0$, for all $\lambda \in \Lambda$, and iii) a uniform covering condition

(Czaja and Li, 2017, Definition 2.1 (b)) which says that the filters' Fourier transform support sets can be covered by a union of finitely many balls. The major differences between (Czaja and Li, 2017) and our results are as follows: (i) the results in (Czaja and Li, 2017) apply exclusively to filters satisfying the uniform covering condition such as, e.g., Weyl-Heisenberg filters with a band-limited prototype function (Czaja and Li, 2017, Proposition 2.3), but do not apply to multi-scale filters such as wavelets, (α)-curvelets, shearlets, and ridgelets (see (Czaja and Li, 2017, Remark 2.2 (b))), (ii) (4.8) as established in (Czaja and Li, 2017) leaves the decay factor $a > 1$ unspecified, whereas our results in (4.28) and (4.31) make the decay factor a explicit (namely, $a = \frac{r^2+1}{r^2-1}$, $r > 1$, in the wavelet case and $a = \frac{1}{2} + \frac{\delta}{R}$, $R \leqslant 2\delta$, in the Weyl-Heisenberg case), (iii) the exponential energy decay result in (4.8) as established in (Czaja and Li, 2017) applies to all $f \in L^2(\mathbb{R}^d)$ and thus, in particular, to Sobolev input signals (owing to $H^s(\mathbb{R}^d) \subseteq L^2(\mathbb{R}^d)$, for all $s > 0$), whereas our decay results in (4.19), (4.28), and (4.31) pertain to Sobolev input signals $f \in H^s(\mathbb{R}^d)$, $s > 0$, only, (iv) the technical elements employed to arrive at the upper bounds on $W_N(f)$, specifically, while the proof in (Czaja and Li, 2017) makes explicit use of the uniform covering property of the filters, our proof of (4.18) is completely oblivious to the (algebraic) structure of the filters, (v) the assumptions imposed on the filters, i.e., the vanishing moments and uniform covering condition in (Czaja and Li, 2017, Definition 2.1 (a)-(b)), in contrast to our Assumption 1, which is less restrictive, and thereby makes our results in Theorem 3 in Section 4.3.1 apply to general (possibly unstructured) filters that, in addition, can be different in different network layers.

4.4. NUMBER OF LAYERS NEEDED

DCNNs used in practice employ potentially hundreds of layers (He et al., 2015). Such network depths entail formidable computational challenges both in training and in operating the network. It is therefore important to understand how many layers are needed to have most of

the input signal energy be contained in the feature vector. This will be done by considering Parseval frames in all layers, i.e., frames with frame bounds $A_n = B_n = 1$, $n \in \mathbb{N}$. The energy conservation result (4.21) then implies that the infinite-depth feature vector $\Phi_\Omega(f) = \bigcup_{n=0}^{\infty} \Phi_\Omega^n(f)$ contains the entire input signal energy according to

$$|||\Phi_\Omega(f)|||^2 = \sum_{n=0}^{\infty} |||\Phi_\Omega^n(f)|||^2 = \|f\|_2^2.$$

Now, the decomposition (4.23) reveals that thanks to $\lim_{N\to\infty} W_N(f) \to 0$, increasing the network depth N implies that the feature vector $\bigcup_{n=0}^{N} \Phi_\Omega^n(f)$ progressively contains a larger fraction of the input signal energy. We formalize the question on the number of layers needed by asking for bounds of the form

$$(1-\varepsilon) \leqslant \frac{\sum_{n=0}^{N} |||\Phi_\Omega^n(f)|||^2}{\|f\|_2^2} \leqslant 1, \tag{4.32}$$

i.e., by determining the network depth N guaranteeing that at least $((1-\varepsilon)\cdot 100)\%$ of the input signal energy are captured by the corresponding depth-N feature vector $\bigcup_{n=0}^{N} \Phi_\Omega^n(f)$. Moreover, (4.32) ensures that the depth-N feature extractor $\bigcup_{n=0}^{N} \Phi_\Omega^n$ exhibits a trivial null-set.

4.4.1. Estimates for band-limited functions

The following results establish handy estimates of the number of layers needed to guarantee (4.32). For pedagogical reasons, we start with the case of band-limited input signals and then proceed in Section 4.4.2 to a more general statement that pertains to Sobolev functions $H^s(\mathbb{R}^d)$.

Corollary 2.

i) Let Ω be the module-sequence (4.1) with filters $\{g_{\lambda_n}\}_{\lambda_n \in \Lambda_n}$ satisfying the conditions in Assumption 1, and let the corresponding frame bounds be $A_n = B_n = 1$, $n \in \mathbb{N}$. Let $\delta > 0$ be the radius of

	$(1-\varepsilon)$					
	0.25	0.5	0.75	0.9	0.95	0.99
wavelets	2	3	4	6	8	11
Weyl-Heisenberg filters	2	4	5	8	10	14
general filters	2	3	7	19	39	199

Table 4.1: Number N of layers needed to ensure that $((1-\varepsilon)\cdot 100)\%$ of the input signal energy is contained in the features generated in the first N network layers.

the spectral gap $B_\delta(0)$ left by the filters $\{g_{\lambda_n}\}_{\lambda_n\in\Lambda_n}$ according to (4.15)*. Furthermore, let $l > \lfloor d/2\rfloor + 1$, $\varepsilon \in (0,1)$, α as defined in* (4.17)*, and $f \in L^2(\mathbb{R}^d)$ L-band-limited. If*

$$N \geqslant \left\lceil \left(\frac{L}{(1-(1-\varepsilon)^{\frac{1}{2l}})\delta} \right)^{1/\alpha} - 1 \right\rceil, \tag{4.33}$$

then (4.32) *holds.*

ii) Assume that the conditions in Theorem 4 i) and ii) hold. For the wavelet case, let $a > 1$ as defined in (4.26) *and $\delta = 1$ (where δ corresponds to the radius of the spectral gap left by the wavelets $\{g_j\}_{j\in\mathbb{Z}\setminus\{0\}}$). For the Weyl-Heisenberg case, let $a > 1$ as defined in* (4.29) *and $\delta \geqslant \frac{R}{2}$ (here, δ corresponds to the radius of the spectral gap left by the Weyl-Heisenberg filters $\{g_k\}_{k\in\mathbb{Z}\setminus\{0\}}$). Moreover, let $l > 1$, $\varepsilon \in (0,1)$, and $f \in L^2(\mathbb{R})$ L-band-limited. If*

$$N \geqslant \left\lceil \log_a\left(\frac{L}{(1-(1-\varepsilon)^{\frac{1}{2l}})\delta} \right) \right\rceil, \tag{4.34}$$

then (4.32) *holds in both cases.*

Proof. The proof is given in Section 4.7.7. □

Corollary 2 nicely shows how the description complexity of the signal class under consideration, namely the bandwidth L and the

dimension d through the decay exponent α defined in (4.17), determine the number N of layers needed to guarantee (4.32). Specifically, (4.33) and (4.34) show that larger bandwidths L and large dimension d render the input signal f more "complex", which requires deeper networks to capture most of the energy of f. The dependence of the lower bounds in (4.33) and (4.34) on the network properties (i.e., the module-sequence Ω) is through the radius δ of the spectral gap left by the filters $\{g_{\lambda_n}\}_{\lambda_n \in \Lambda_n}$ and the decay factor a.

The following numerical example provides quantitative insights on the influence of the parameter ε on (4.33) and (4.34). Specifically, we set $L = 1$, $d = 1$ (which implies $\alpha = 1$, see (4.17)), $r = 2$ (which implies $a = \frac{5}{3}$ in the wavelet case, see (4.26)), $R = \delta = 1$ (which implies $a = \frac{3}{2}$ in the Weyl-Heisenberg case, see (4.29)), $l = 1.0001$, and show in Table 4.1 the number N of layers needed according to (4.33) and (4.34) for different values of ε. The results show that 95% of the input signal energy are contained in the first 8 layers in the wavelet case and the first 10 layers in the Weyl-Heisenberg case. We can therefore conclude that in practice a relatively small number of layers is needed to have most of the input signal energy be contained in the feature vector. In contrast, for general filters, where we can guarantee polynomial energy decay only, at least $N = 39$ layers are needed to absorb 95% of the input signal energy. We hasten to add, however, that (4.18) simply *guarantees* polynomial energy decay and therefore does not preclude the energy from decaying faster in practice.

4.4.2. Estimates for Sobolev functions

We proceed with the estimates on the number of layers for Sobolev-class input signals.

Corollary 3.

i) Let Ω be the module-sequence (4.1) with filters $\{g_{\lambda_n}\}_{\lambda_n \in \Lambda_n}$ satisfying the conditions in Assumption 1, and let the corresponding frame bounds be $A_n = B_n = 1$, $n \in \mathbb{N}$. Let $\delta > 0$ be the radius of the spectral gap $B_\delta(0)$ left by the filters $\{g_{\lambda_n}\}_{\lambda_n \in \Lambda_n}$ according to

(4.15). *Furthermore, let* $l > \lfloor d/2 \rfloor + 1$, $\varepsilon \in (0,1)$, α *as defined in* (4.17), *and* $f \in H^s(\mathbb{R}^d) \backslash \{0\}$, *for* $s > 0$. *If*

$$N \geqslant \left\lceil \left(\frac{2l \, \|f\|_{H^s}^{2/\gamma}}{\varepsilon^{1/\gamma} \, \delta \|f\|_2^{2/\gamma}} \right)^{1/\alpha} - 1 \right\rceil, \tag{4.35}$$

where $\gamma := \min\{1, 2s\}$, *then* (4.32) *holds.*

ii) Assume that the conditions in Theorem 4 i) and ii) hold. For the wavelet case, let $a > 1$ *as defined in* (4.26) *and* $\delta = 1$ $\big($*where* δ *corresponds to the radius of the spectral gap left by the wavelets* $\{g_j\}_{j \in \mathbb{Z} \backslash \{0\}}\big)$. *For the Weyl-Heisenberg case, let* $a > 1$ *as defined in* (4.29) *and* $\delta \geqslant \frac{R}{2}$ $\big($*here,* δ *corresponds to the radius of the spectral gap left by the Weyl-Heisenberg filters* $\{g_k\}_{k \in \mathbb{Z} \backslash \{0\}}\big)$. *Furthermore, let* $l > 1$, $\varepsilon \in (0,1)$, *and* $f \in H^s(\mathbb{R}) \backslash \{0\}$, *for* $s > 0$. *If*

$$N \geqslant \left\lceil \log_a \left(\frac{2l \, \|f\|_{H^s}^{2/\gamma}}{\varepsilon^{1/\gamma} \, \delta \|f\|_2^{2/\gamma}} \right) \right\rceil, \tag{4.36}$$

where $\gamma := \min\{1, 2s\}$, *then* (4.32) *holds in both cases.*

Proof. The proof is given in Section 4.7.8. □

As already mentioned in Section 4.2, Sobolev spaces $H^s(\mathbb{R}^d)$ contain a wide range of practically relevant signal classes. The results in Corollary 3 therefore provide—for a wide variety of input signals—a picture of how many layers are needed to have most of the input signal energy be contained in the feature vector.

The width of the networks considered throughout this thesis is, in principle, infinite as the sets Λ_n need to be countably infinite in order to guarantee that the frame property (4.2) is satisfied. For input signals that are essentially band-limited, the number of "operationally significant nodes" will, however, be finite in practice. For a treatment of this aspect as well as results on depth-width tradeoffs, the interested reader is referred to (Wiatowski et al., 2017).

4.5. DEPTH-CONSTRAINED NETWORKS

We now turn to the design of scattering networks of fixed (possibly small) depth N that capture most of the input signal's energy. This will be formalized by seeking wavelet and Weyl-Heisenberg filters that, for given $\varepsilon > 0$ and given depth $N \in \mathbb{N}$, result in feature extractors satisfying

$$(1-\varepsilon)\|f\|_2^2 \leqslant \sum_{n=0}^{N} |||\Phi_\Omega^n(f)|||^2 \leqslant \|f\|_2^2, \quad \forall f \in L^2(\mathbb{R}). \tag{4.37}$$

The next result explains how to choose r in the wavelet and R in the Weyl-Heisenberg case so as to satisfy (4.37). In particular, we shall see that for every (possibly small) $\varepsilon > 0$ and every $N \in \mathbb{N}$, say $\varepsilon = 0.01$ and $N = 1$, there exist $r > 1$ and $R > 0$ such that (4.37) holds.

Corollary 4. *Assume that the conditions in Theorem 4 i) and ii) hold. For the wavelet case, let $r > 1$ and $\delta = 1$ $\big($where δ corresponds to the radius of the spectral gap left by the wavelets $\{g_j\}_{j\in\mathbb{Z}\setminus\{0\}}\big)$. For the Weyl-Heisenberg case, let $R > 0$, $\delta \geqslant \frac{R}{2}$ $\big($here, δ corresponds to the radius of the spectral gap left by the Weyl-Heisenberg filters $\{g_k\}_{k\in\mathbb{Z}\setminus\{0\}}\big)$. Moreover, take $f \in H^s(\mathbb{R})\setminus\{0\}$, $s > 0$, fix $\varepsilon \in (0,1)$ and $N \in \mathbb{N}$, let $l > \frac{1}{2}\,\varepsilon^{1/\gamma}\,\delta$, where $\gamma := \min\{1, 2s\}$, and define*

$$\kappa := \left(\frac{2l\,\|f\|_{H^s}^{2/\gamma}}{\varepsilon^{1/\gamma}\delta\|f\|_2^{2/\gamma}}\right)^{1/N}.$$

If, in the wavelet case,

$$1 < r \leqslant \sqrt{\frac{\kappa+1}{\kappa-1}}, \tag{4.38}$$

or, in the Weyl-Heisenberg case,

$$0 < R \leqslant \frac{\delta}{\kappa - \frac{1}{2}}, \tag{4.39}$$

then (4.37) *holds.*

Proof. The proof is given in Section 4.7.9. □

4.6. A FEATURE EXTRACTOR WITH A NON-TRIVIAL NULL-SET

In this section, we show, by way of example, that employing filters Ψ_n which satisfy the frame property (4.2) alone does *not* guarantee that the feature extractor Φ_Ω defined in (3.12) satisfies

$$A_\Omega \|f\|_2^2 \leqslant |||\Phi_\Omega(f)|||^2, \qquad \forall f \in L^2(\mathbb{R}^d),$$

for some $A_\Omega > 0$. The existence of such a lower bound $A_\Omega > 0$ would imply a trivial null-set for the feature extractor Φ_Ω and thereby ensure that the only signal f that maps to the all-zeros feature vector is $f = 0$.

Our example employs, in every network layer, filters $\Psi = \{\chi\} \cup \{g_k\}_{k\in\mathbb{Z}}$ that satisfy the Littlewood-Paley condition (4.16) with $A_n = B_n = 1$, $n \in \mathbb{N}$, and where g_0 is such that $\widehat{g_0}(\omega) = 1$, for $\omega \in \overline{B_1(0)}$. We emphasize that no further restrictions are imposed on the filters $\{\chi\} \cup \{g_k\}_{k\in\mathbb{Z}}$, specifically χ need not be of low-pass nature and the filters $\{g_k\}_{k\in\mathbb{Z}}$ may be structured (such as wavelets, see Sections 2.2.1 and 2.2.2) or unstructured (such as random filters (Ranzato et al., 2007; Jarrett et al., 2009)), as long as they satisfy the Littlewood-Paley condition (4.16). Now, consider the input signal $f \in L^2(\mathbb{R}^d)$ according to

$$\widehat{f}(\omega) := (1 - |\omega|)_+^l, \quad \omega \in \mathbb{R}^d,$$

with $l > \lfloor d/2 \rfloor + 1$. Then $f * g_0 = f$, owing to $\operatorname{supp}(\widehat{f}) = \overline{B_1(0)}$ and $\widehat{g_0}(\omega) = 1$, for $\omega \in \overline{B_1(0)}$. Moreover, $\widehat{f}$ is a positive definite radial basis function (Wendland, 2004, Theorem 6.20) and hence by (Wendland, 2004, Theorem 6.18) $f(x) \geqslant 0$, $x \in \mathbb{R}^d$, which, in turn, implies $|f| = f$. This yields

$$U[q_0^N]f = \big|\cdots\big||f * g_0| * g_0\big|\cdots * g_0\big| = f,$$

for $q_0^N := (0, 0, \ldots, 0) \in \mathbb{Z}^N$ and $N \in \mathbb{N}$. Owing to the energy decomposition identity (4.22), together with $A_\Omega^N = B_\Omega^N = 1$, $N \in \mathbb{N}$,

which, in turn, is by $A_n = B_n = 1$, $n \in \mathbb{N}$, we have

$$\begin{aligned}\|f\|_2^2 &= \sum_{n=0}^{N-1} |||\Phi_\Omega^n(f)|||^2 + W_N(f) \\ &= \sum_{n=0}^{N-1} |||\Phi_\Omega^n(f)|||^2 + \underbrace{\|U[q_0^N]f\|_2^2}_{=\|f\|_2^2} + \sum_{q \in \mathbb{Z}^N \setminus \{q_0^N\}} \|U[q]f\|_2^2,\end{aligned}$$

for $N \in \mathbb{N}$. This implies

$$\sum_{n=0}^{N-1} |||\Phi_\Omega^n(f)|||^2 + \sum_{q \in \mathbb{Z}^N \setminus \{q_0^N\}} \|U[q]f\|_2^2 = 0. \tag{4.40}$$

As both terms in (4.40) are positive, we can conclude that $\sum_{n=0}^{N-1} |||\Phi_\Omega^n(f)|||^2 = 0$, $N \in \mathbb{N}$, and thus

$$|||\Phi_\Omega(f)|||^2 = \sum_{n=0}^{\infty} |||\Phi_\Omega^n(f)|||^2 = 0.$$

Since $|||\Phi_\Omega(f)|||^2 = 0$ implies $\Phi_\Omega(f) = 0$, we have constructed a non-zero f, namely

$$f(x) = \int_{\mathbb{R}^d} (1 - |\omega|)_+^l e^{\,2\pi i \langle x,\, \omega \rangle} \mathrm{d}\omega,$$

that maps to the all-zeros feature vector, i.e., $f \in \mathcal{N}(\Phi_\Omega)$.

The point of this example is the following. Owing to the nature of $\widehat{g_0}(\omega)$ (namely, $\widehat{g_0}(\omega) = 1$, for $\omega \in \overline{B_1(0)}$) and the Littlewood-Paley condition

$$|\widehat{\chi}(\omega)|^2 + \sum_{k \in \mathbb{Z}} |\widehat{g_k}(\omega)|^2 = 1, \quad \text{a.e. } \omega \in \mathbb{R}^d,$$

it follows that neither the output-generating filter χ nor any of the other filters g_k, $k \in \mathbb{Z}\backslash\{0\}$, can have spectral support in $\overline{B_1(0)}$. Consequently, the only non-zero contribution to the feature vector can come from

$$U[q_0^N]f = f,$$

which, however, thanks to $\operatorname{supp}(\widehat{f}\,) = \overline{B_1(0)}$, is spectrally disjoint from the output-generating filter χ. Therefore, $\Phi_\Omega(f)$ will be identically equal to 0. Assumption 1 disallows this situation as it forces the filters g_k, $k \in \mathbb{Z}$, to be of high-pass nature which, in turn, implies that χ must have low-pass characteristics. The punch-line of our general results on energy conservation, be it for finite N or for $N \to \infty$, is that Assumption 1 in combination with the frame property and the modulus non-linearity prohibit a non-trivial null-set *in general*.

4.7. PROOFS

4.7.1. Proof of Lemma 5

Cartoon functions, introduced in (Donoho, 2001), satisfy mild decay properties and are piecewise continuously differentiable apart from curved discontinuities along C^2-hypersurfaces (for a formal definition we refer to Definition 7 in Section 3.4.3). Even though cartoon functions are in general discontinuous, they still admit Sobolev regularity. The following result formalizes this statement.

Lemma 5. *Let $K > 0$. Then $\mathcal{C}^K_{\mathrm{CART}} \subseteq H^s(\mathbb{R}^d)$, for all $s \in (0, 1/2)$.*

Proof. Let $(f_1 + \mathbb{1}_B f_2) \in \mathcal{C}^K_{\mathrm{CART}}$. We first establish $\in H^s(\mathbb{R}^d)$, for all $s \in (0, 1/2)$. To this end, we define the Sobolev-Slobodeckij semi-norm (Runst and Sickel, 1996, Section 2.1.2)

$$|f|_{H^s} := \Big(\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \frac{|f(x) - f(y)|^2}{|x-y|^{2s+d}} \mathrm{d}x \, \mathrm{d}y \Big)^{1/s},$$

and note that, thanks to (Runst and Sickel, 1996, Section 2.1.2), $\mathbb{1}_B \in H^s(\mathbb{R}^d)$ if $|\mathbb{1}_B|_{H^s} < \infty$. We have

$$\begin{aligned} |\mathbb{1}_B|^s_{H^s} &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \frac{|\mathbb{1}_B(x) - \mathbb{1}_B(y)|^2}{|x-y|^{2s+d}} \mathrm{d}x \, \mathrm{d}y \\ &= \int_{\mathbb{R}^d} \frac{1}{|t|^{2s+d}} \int_{\mathbb{R}^d} |\mathbb{1}_B(x) - \mathbb{1}_B(x-t)|^2 \mathrm{d}x \, \mathrm{d}t, \end{aligned}$$

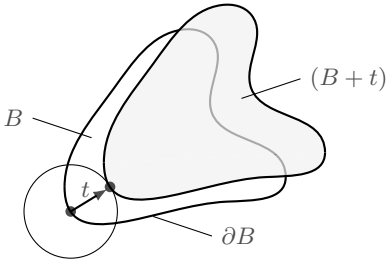

Fig. 4.8: Illustration in dimension $d = 2$. The set $(B + t)$ (grey) is obtained by translating the set B (white) by $t \in \mathbb{R}^2$. The symmetric difference $B\Delta(B + t)$ is contained in $(\partial B + B_{|t|}(0))$, the tube of radius $|t|$ around the boundary ∂B of B.

where we employed the change of variables $t = x - y$. Next, we note that, for fixed $t \in \mathbb{R}^d$, the function

$$h_t(x) := |\mathbb{1}_B(x) - \mathbb{1}_B(x - t)|^2$$

satisfies $h_t(x) = 1$, for $x \in S_t$, where

$$\begin{aligned} S_t := &\{x \in \mathbb{R}^d \,|\, x \in B \;\text{ and }\; x - t \notin B\} \\ &\cup \{x \in \mathbb{R}^d \,|\, x \notin B \;\text{ and }\; x - t \in B\} = B\Delta(B + t), \end{aligned} \tag{4.41}$$

and $h_t(x) = 0$, for $x \in \mathbb{R}^d \backslash S_t$. It follows from (4.41) that

$$\mathrm{vol}^d(S_t) \leqslant 2\,\mathrm{vol}^d(B), \quad \forall\, t \in \mathbb{R}^d. \tag{4.42}$$

Moreover, owing to $S_t \subseteq \big(\partial B + B_{|t|}(0)\big)$, where $(\partial B + B_{|t|}(0))$ is a tube of radius $|t|$ around the boundary ∂B of B (see Fig. 4.8), and Lemma 4 in Section 3.6.6, there exists a constant $C_{\partial B} > 0$ such that

$$\mathrm{vol}^d(S_t) \leqslant \mathrm{vol}^d(\partial B + B_{|t|}(0)) \leqslant C_{\partial B}|t|, \tag{4.43}$$

for all $t \in \mathbb{R}^d$ with $|t| \leqslant 1$. Next, fix R such that $0 < R < 1$. Then,

$$|\mathbb{1}_B|^s_{H^s} = \int_{\mathbb{R}^d} \frac{1}{|t|^{2s+d}} \int_{\mathbb{R}^d} |\mathbb{1}_B(x) - \mathbb{1}_B(x - t)|^2 \mathrm{d}x\, \mathrm{d}t$$

$$
\begin{aligned}
&= \int_{\mathbb{R}^d} \frac{1}{|t|^{2s+d}} \int_{\mathbb{R}^d} h_t(x) \mathrm{d}x \, \mathrm{d}t \\
&= \int_{\mathbb{R}^d} \frac{1}{|t|^{2s+d}} \int_{S_t} 1 \mathrm{d}x \, \mathrm{d}t = \int_{\mathbb{R}^d} \frac{\mathrm{vol}^d(S_t)}{|t|^{2s+d}} \mathrm{d}t \\
&\leqslant \int_{\mathbb{R}^d \setminus B_R(0)} \frac{2 \, \mathrm{vol}^d(B)}{|t|^{2s+d}} \mathrm{d}t + \int_{B_R(0)} \frac{C_{\partial B}}{|t|^{2s+d-1}} \mathrm{d}t \qquad (4.44) \\
&= 2 \, \mathrm{vol}^d(B) \, \mathrm{vol}^{d-1}(\partial B_1(0)) \underbrace{\int_R^\infty r^{-(2s+1)} \mathrm{d}r}_{=: I_1} \\
&+ C_{\partial B} \, \mathrm{vol}^{d-1}(\partial B_1(0)) \underbrace{\int_0^R r^{-2s} \mathrm{d}r}_{=: I_2}, \qquad (4.45)
\end{aligned}
$$

where in (4.44) we employed (4.42) and (4.43), and in the last step we introduced polar coordinates. The integral I_1 is finite for all $s > 0$, while I_2 is finite for all $s < 1/2$. Moreover, $\mathrm{vol}^d(B) = \int_B 1 \mathrm{d}x$ is finite owing to B being compact. We can therefore conclude that (4.45) is finite for $s \in (0, 1/2)$, and hence $\mathbb{1}_B \in H^s(\mathbb{R}^d)$, for $s \in (0, 1/2)$. To see that $(f_1 + \mathbb{1}_B f_2) \in H^s(\mathbb{R}^d)$, for $s \in (0, 1/2)$, we first note that

$$
|f_1 + \mathbb{1}_B f_2|_{H^s} \leqslant |f_1|_{H^s} + |\mathbb{1}_B f_2|_{H^s}, \qquad (4.46)
$$

which is thanks to the sub-additivity of the semi-norm $|\cdot|_{H^s}$. Now, the first term on the RHS of (4.46) is finite owing to $f_1 \in H^{1/2}(\mathbb{R}^d) \subseteq H^s(\mathbb{R}^d)$, for all $s \in (0, 1/2)$. For the second term on the RHS, we start by noting that

$$
|\mathbb{1}_B f_2|_{H^s}^s = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \frac{|(\mathbb{1}_B f_2)(x) - (\mathbb{1}_B f_2)(y)|^2}{|x - y|^{2s+d}} \mathrm{d}x \, \mathrm{d}y \qquad (4.47)
$$

and

$$
\begin{aligned}
&|(\mathbb{1}_B f_2)(x) - (\mathbb{1}_B f_2)(y)|^2 \\
&= |(\mathbb{1}_B(x) - \mathbb{1}_B(y)) f_2(x) + (f_2(x) - f_2(y)) \mathbb{1}_B(y)|^2 \\
&\leqslant 2 |(\mathbb{1}_B(x) - \mathbb{1}_B(y))|^2 |f_2(x)|^2 \qquad (4.48) \\
&+ 2 |(f_2(x) - f_2(y))|^2 |\mathbb{1}_B(y)|^2, \qquad (4.49)
\end{aligned}
$$

where (4.48) and (4.49) are thanks to $|a+b|^2 \leqslant 2|a|^2 + 2|b|^2$, for $a, b \in \mathbb{C}$. Substituting (4.48) and (4.49) into (4.47) and noting that $|f_2(x)|^2 \leqslant \|f_2\|_\infty^2 \leqslant K^2$, $x \in \mathbb{R}^d$, which is by assumption, and $\mathbb{1}_B(y) \leqslant 1$, $y \in \mathbb{R}^d$, implies

$$|\mathbb{1}_B f_2|_{H^s}^s \leqslant 2K^2 |\mathbb{1}_B|_{H^s}^s + 2|f_2|_{H^s}^s < \infty, \tag{4.50}$$

where in the last step we used $\mathbb{1}_B \in H^s(\mathbb{R}^d)$, established above, and $f_2 \in H^{1/2}(\mathbb{R}^d) \subseteq H^s(\mathbb{R}^d)$, both for all $s \in (0, 1/2)$. This completes the proof. □

4.7.2. Proof of statement i) in Theorem 3

We start by establishing (4.18) with $\alpha = \log_2(\sqrt{d/(d-1/2)})$, for all $d \geqslant 1$. Then, we sharpen our result in the 1-D case by proving that (4.18) holds for $d = 1$ with $\alpha = 1$. This leads to a significant improvement, in the 1-D case, of the decay exponent from $\log_2(\sqrt{d/(d-1/2)}) = \frac{1}{2}$ to 1.

The idea for the proof of (4.18) for $\alpha = \log_2(\sqrt{d/(d-1/2)})$, for all $d \geqslant 1$, is to establish that[2]

$$\sum_{q \in \Lambda_n \times \Lambda_{n+1} \times \cdots \times \Lambda_{n+N-1}} \|U[q]f\|_2^2$$
$$\leqslant C_n^{n+N-1} \int_{\mathbb{R}^d} \left|\widehat{f}(\omega)\right|^2 \Big(1 - \Big|\widehat{r}_l\Big(\frac{\omega}{N^\alpha \delta}\Big)\Big|^2\Big) \mathrm{d}\omega, \quad \forall N \in \mathbb{N}, \tag{4.51}$$

where

$$C_n^{n+N-1} := \prod_{k=n}^{n+N-1} \max\{1, B_k\}.$$

Setting $n = 1$ in (4.51) and noting that $C_1^N = B_\Omega^N$ yields the desired result (4.18). We proceed by induction over the path length $\ell(q) := N$,

[2] We prove the more general result (4.51) for technical reasons, concretely in order to be able to argue by induction over path lengths with flexible starting index n.

for $q = (\lambda_n, \lambda_{n+1}, \ldots, \lambda_{n+N-1}) \in \Lambda_n \times \Lambda_{n+1} \times \cdots \times \Lambda_{n+N-1}$. Starting with the base case $N = 1$, we have

$$\begin{aligned}
\sum_{q \in \Lambda_n} \|U[q]f\|_2^2 &= \sum_{\lambda_n \in \Lambda_n} \|f * g_{\lambda_n}\|_2^2 \\
&= \int_{\mathbb{R}^d} \sum_{\lambda_n \in \Lambda_n} |\widehat{g_{\lambda_n}}(\omega)|^2 |\widehat{f}(\omega)|^2 \mathrm{d}\omega && (4.52) \\
&\leqslant B_n \int_{\mathbb{R}^d \setminus B_\delta(0)} |\widehat{f}(\omega)|^2 \mathrm{d}\omega && (4.53) \\
&\leqslant \underbrace{\max\{1, B_n\}}_{=C_n^n} \int_{\mathbb{R}^d} \left|\widehat{f}(\omega)\right|^2 \Big(1 - \Big|\widehat{r_l}\Big(\frac{\omega}{\delta}\Big)\Big|^2\Big) \mathrm{d}\omega, && (4.54)
\end{aligned}$$

for all $n \in \mathbb{N}$, where (4.52) is by Parseval's formula, (4.53) is thanks to (4.15) and (4.16), and (4.54) is due to $\mathrm{supp}(\widehat{r_l}) \subseteq B_1(0)$ and $0 \leqslant \widehat{r_l}(\omega) \leqslant 1$, for $\omega \in \mathbb{R}^d$. The inductive step is established as follows. Let $N > 1$ and suppose that (4.51) holds for all paths q of length $\ell(q) = N - 1$, i.e.,

$$\begin{aligned}
&\sum_{q \in \Lambda_n \times \Lambda_{n+1} \times \cdots \times \Lambda_{n+N-2}} \|U[q]f\|_2^2 \\
&\leqslant C_n^{n+N-2} \int_{\mathbb{R}^d} \left|\widehat{f}(\omega)\right|^2 \Big(1 - \Big|\widehat{r_l}\Big(\frac{\omega}{(N-1)^\alpha \delta}\Big)\Big|^2\Big) \mathrm{d}\omega, && (4.55)
\end{aligned}$$

for all $n \in \mathbb{N}$. We start by noting that every path $\tilde{q} \in \Lambda_n \times \Lambda_{n+1} \times \ldots \times \Lambda_{n+N-1}$ of length $\ell(\tilde{q}) = N$, with arbitrary starting index n, can be decomposed into a path $q \in \Lambda_{n+1} \times \ldots \times \Lambda_{n+N-1}$ of length $\ell(q) = N - 1$ and an index $\lambda_n \in \Lambda_n$ according to $\tilde{q} = (\lambda_n, q)$. Thanks to (4.3) we have $U[\tilde{q}] = U[(\lambda_n, q)] = U[q]U_n[\lambda_n]$, which yields

$$\begin{aligned}
&\sum_{q \in \Lambda_n \times \Lambda_{n+1} \times \cdots \times \Lambda_{n+N-1}} \|U[q]f\|_2^2 \\
&= \sum_{\lambda_n \in \Lambda_n} \sum_{q \in \Lambda_{n+1} \times \cdots \times \Lambda_{n+N-1}} \|U[q]\big(U_n[\lambda_n]f\big)\|_2^2, && (4.56)
\end{aligned}$$

for all $n \in \mathbb{N}$. We proceed by examining the inner sum on the RHS of (4.56). Invoking the induction hypothesis (4.55) with n replaced by

$(n+1)$ and employing Parseval's formula, we get

$$\begin{aligned}
&\sum_{q\in\Lambda_{n+1}\times\cdots\times\Lambda_{n+N-1}} \|U[q]\big(U_n[\lambda_n]f\big)\|_2^2 \\
&\leqslant C_{n+1}^{n+N-1}\int_{\mathbb{R}^d} \big|\widehat{U_n[\lambda_n]f}(\omega)\big|^2\Big(1-\Big|\widehat{r}_l\Big(\frac{\omega}{(N-1)^\alpha\delta}\Big)\Big|^2\Big)\mathrm{d}\omega \\
&= C_{n+1}^{n+N-1}\big(\|U_n[\lambda_n]f\|_2^2 - \|(U_n[\lambda_n]f)*r_{l,N-1,\alpha,\delta}\|_2^2\big) \\
&= C_{n+1}^{n+N-1}\big(\|f*g_{\lambda_n}\|_2^2 - \||f*g_{\lambda_n}|*r_{l,N-1,\alpha,\delta}\|_2^2\big), \qquad (4.57)
\end{aligned}$$

for $n\in\mathbb{N}$, where $r_{l,N-1,\alpha,\delta}$ is the inverse Fourier transform of $\widehat{r}_l\big(\frac{\omega}{(N-1)^\alpha\delta}\big)$. Next, we note that $\widehat{r}_l\big(\frac{\omega}{(N-1)^\alpha\delta}\big)$ is a positive definite radial basis function (Wendland, 2004, Theorem 6.20) and hence by (Wendland, 2004, Theorem 6.18) $r_{l,N-1,\alpha,\delta}(x)\geqslant 0$, for $x\in\mathbb{R}^d$. Furthermore, it follows from Lemma 6, stated below, that for all $\{\nu_{\lambda_n}\}_{\lambda_n\in\Lambda_n}\subseteq\mathbb{R}^d$, we have

$$\||f*g_{\lambda_n}|*r_{l,N-1,\alpha,\delta}\|_2^2 \geqslant \|f*g_{\lambda_n}*(M_{\nu_{\lambda_n}}r_{l,N-1,\alpha,\delta})\|_2^2. \qquad (4.58)$$

Here, we note that choosing the modulation factors $\{\nu_{\lambda_n}\}_{\lambda_n\in\Lambda_n}\subseteq\mathbb{R}^d$ appropriately (see (4.62) below) will be key to establishing the inductive step.

Lemma 6. *(Mallat, 2012, Lemma 2.7) Let $f,g\in L^2(\mathbb{R}^d)$ with $g(x)\geqslant 0$, for $x\in\mathbb{R}^d$. Then,*

$$\||f|*g\|_2^2\geqslant\|f*(M_\omega g)\|_2^2,\quad\forall\omega\in\mathbb{R}^d.$$

Inserting (4.57) and (4.58) into the inner sum on the RHS of (4.56) yields

$$\begin{aligned}
&\sum_{q\in\Lambda_n\times\Lambda_{n+1}\times\cdots\times\Lambda_{n+N-1}}\|U[q]f\|_2^2 \\
&\leqslant C_{n+1}^{n+N-1}\sum_{\lambda_n\in\Lambda_n}\Big(\|f*g_{\lambda_n}\|_2^2-\|f*g_{\lambda_n}*(M_{\nu_{\lambda_n}}r_{l,N-1,\alpha,\delta})\|_2^2\Big) \\
&= C_{n+1}^{n+N-1}\int_{\mathbb{R}^d}|\widehat{f}(\omega)|^2h_{n,N,\alpha,\delta}(\omega)\mathrm{d}\omega,\quad\forall N\in\mathbb{N}, \qquad (4.59)
\end{aligned}$$

where we applied Parseval's formula together with $\widehat{M_\omega f} = T_\omega \widehat{f}$, for $f \in L^2(\mathbb{R}^d)$, and $\omega \in \mathbb{R}^d$ and set

$$h_{n,N,\alpha,\delta}(\omega) := \sum_{\lambda_n \in \Lambda_n} |\widehat{g_{\lambda_n}}(\omega)|^2 \Big(1 - \Big|\widehat{r_l}\Big(\frac{\omega - \nu_{\lambda_n}}{(N-1)^\alpha \delta}\Big)\Big|^2\Big). \tag{4.60}$$

The key step is now to establish—by appropriately choosing $\{\nu_{\lambda_n}\}_{\lambda_n \in \Lambda_n} \subseteq \mathbb{R}^d$—the upper bound

$$h_{n,N,\alpha,\delta}(\omega) \leqslant \max\{1, B_n\}\Big(1 - \Big|\widehat{r_l}\Big(\frac{\omega}{N^\alpha \delta}\Big)\Big|^2\Big), \quad \forall\, \omega \in \mathbb{R}^d, \tag{4.61}$$

which upon noting that $C_n^{n+N-1} = \max\{1, B_n\}\, C_{n+1}^{n+N-1}$ yields (4.51) and thereby completes the proof. We start by defining $H_{A_{\lambda_n}}$, for $\lambda_n \in \Lambda_n$, to be the orthant supporting $\widehat{g_{\lambda_n}}$, i.e., $\mathrm{supp}(\widehat{g_{\lambda_n}}) \subseteq H_{A_{\lambda_n}}$, where $A_{\lambda_n} \in O(d)$ (see Assumption 1). Furthermore, for $\lambda_n \in \Lambda_n$, we choose the modulation factors according to

$$\nu_{\lambda_n} := A_{\lambda_n} \nu \in \mathbb{R}^d, \tag{4.62}$$

where the components of $\nu \in \mathbb{R}^d$ are given by $\nu_k := (1 + 2^{-1/2})\frac{\delta}{d}$, for $k \in \{1, \ldots, d\}$. Invoking (4.14) and (4.15), we get

$$\begin{aligned} h_{n,N,\alpha,\delta}(\omega) &= \sum_{\lambda_n \in \Lambda_n} |\widehat{g_{\lambda_n}}(\omega)|^2 \Big(1 - \Big|\widehat{r_l}\Big(\frac{\omega - \nu_{\lambda_n}}{(N-1)^\alpha \delta}\Big)\Big|^2\Big) \\ &= \sum_{\lambda_n \in \Lambda_n} |\widehat{g_{\lambda_n}}(\omega)|^2 \mathbb{1}_{S_{\lambda_n,\delta}}(\omega) \Big(1 - \Big|\widehat{r_l}\Big(\frac{\omega - \nu_{\lambda_n}}{(N-1)^\alpha \delta}\Big)\Big|^2\Big), \end{aligned} \tag{4.63}$$

for all $\omega \in \mathbb{R}^d$, where $S_{\lambda_n,\delta} := H_{A_{\lambda_n}} \backslash B_\delta(0)$. For the first canonical orthant $H = \{x \in \mathbb{R}^d \mid x_k \geqslant 0,\ k = 1,\ldots,d\}$ we show in Lemma 7 below that

$$\Big|\widehat{r_l}\Big(\frac{\omega - \nu}{(N-1)^\alpha \delta}\Big)\Big| \geqslant \Big|\widehat{r_l}\Big(\frac{\omega}{N^\alpha \delta}\Big)\Big|, \tag{4.64}$$

for all $\omega \in H \backslash B_\delta(0)$ and all $N \geqslant 2$. This will allow us to deduce

$$\Big|\widehat{r_l}\Big(\frac{\omega - \nu_{\lambda_n}}{(N-1)^\alpha \delta}\Big)\Big| \geqslant \Big|\widehat{r_l}\Big(\frac{\omega}{N^\alpha \delta}\Big)\Big|, \tag{4.65}$$

for all $\omega \in S_{\lambda_n,\delta}$, all $\lambda_n \in \Lambda_n$, and all $N \geqslant 2$, where $S_{\lambda_n,\delta} = H_{A_{\lambda_n}} \backslash B_\delta(0)$, simply by noting that

$$\begin{aligned}
\Big|\widehat{r}_l\Big(\frac{\omega-\nu_{\lambda_n}}{(N-1)^\alpha\delta}\Big)\Big| &= \bigg(1-\Big|\frac{A_{\lambda_n}(\omega'-\nu)}{(N-1)^\alpha\delta}\Big|\bigg)_+^l \\
&= \Big(1-\Big|\frac{\omega'-\nu}{(N-1)^\alpha\delta}\Big|\Big)_+^l = \Big|\widehat{r}_l\Big(\frac{\omega'-\nu}{(N-1)^\alpha\delta}\Big)\Big| && (4.66)\\
&\geqslant \Big|\widehat{r}_l\Big(\frac{\omega'}{N^\alpha\delta}\Big)\Big| = \Big(1-\Big|\frac{\omega'}{N^\alpha\delta}\Big|\Big)_+^l && (4.67)\\
&= \bigg(1-\Big|\frac{A_{\lambda_n}\omega'}{N^\alpha\delta}\Big|\bigg)_+^l = \Big|\widehat{r}_l\Big(\frac{\omega}{N^\alpha\delta}\Big)\Big|, && (4.68)
\end{aligned}$$

for $\omega = A_{\lambda_n}\omega' \in H_{A_{\lambda_n}} \backslash B_\delta(0)$, where $\omega' \in H \backslash B_\delta(0)$. Here, (4.66) and (4.68) are thanks to $|\omega| = |A_{\lambda_n}\omega|$, which is by $A_{\lambda_n} \in O(d)$, and the inequality in (4.67) is due to (4.64). Insertion of (4.65) into (4.63) then yields

$$\begin{aligned}
h_{n,N,\alpha,\delta}(\omega) &\leqslant \sum_{\lambda_n\in\Lambda_n} |\widehat{g_{\lambda_n}}(\omega)|^2 \mathbb{1}_{S_{\lambda_n,\delta}}(\omega)\Big(1-\Big|\widehat{r}_l\Big(\frac{\omega}{N^\alpha\delta}\Big)\Big|^2\Big) \\
&= \sum_{\lambda_n\in\Lambda_n} |\widehat{g_{\lambda_n}}(\omega)|^2\Big(1-\Big|\widehat{r}_l\Big(\frac{\omega}{N^\alpha\delta}\Big)\Big|^2\Big) && (4.69)\\
&\leqslant \max\{1,B_n\}\Big(1-\Big|\widehat{r}_l\Big(\frac{\omega}{N^\alpha\delta}\Big)\Big|^2\Big), \qquad \forall\,\omega\in\mathbb{R}^d, && (4.70)
\end{aligned}$$

where in (4.69) we employed Assumption 1, and (4.70) is thanks to (4.16). This establishes (4.61) and completes the proof of (4.18) for $\alpha = \log_2(\sqrt{d/(d-1/2)})$, for all $d \geqslant 1$.

It remains to show (4.64), which is accomplished through the following lemma.

Lemma 7. *Let $\alpha := \log_2\big(\sqrt{d/(d-1/2)}\big)$, $\widehat{r}_l : \mathbb{R}^d \to \mathbb{R}$, $\widehat{r}_l(\omega) := (1-|\omega|)_+^l$, with $l > \lfloor d/2\rfloor + 1$, and define $\nu \in \mathbb{R}^d$ to have components $\nu_k = (1+2^{-1/2})\frac{\delta}{d}$, for $k \in \{1,\dots,d\}$. Then,*

$$\Big|\widehat{r}_l\Big(\frac{\omega-\nu}{(N-1)^\alpha\delta}\Big)\Big| \geqslant \Big|\widehat{r}_l\Big(\frac{\omega}{N^\alpha\delta}\Big)\Big|, \qquad (4.71)$$

for all $\omega \in H\backslash B_\delta(0)$ and all $N \geqslant 2$.

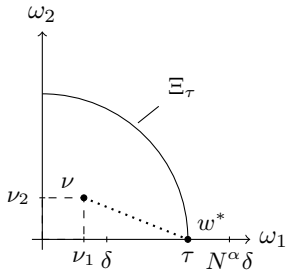

Fig. 4.9: Illustration in dimension $d = 2$. The mapping $\omega \mapsto |\omega - \nu|^2$, $\omega \in \Xi_\tau = \{\omega = (\omega_1, \omega_2) \in \mathbb{R}^2 \mid |\omega| = \tau,\, \omega_1 \geqslant 0,\, \omega_2 \geqslant 0\}$, computes the squared Euclidean distance between an element ω of the spherical segment Ξ_τ and the vector $\nu = (\nu_1, \nu_2)$ with components $\nu_k = (1 + 2^{-1/2})\frac{\delta}{2}$, $k \in \{1, 2\}$. The mapping attains its maxima along the coordinate axes, e.g., for $\omega^* = (\tau, 0) \in \Xi_\tau$.

Proof. The key idea of the proof is to employ a monotonicity argument. Specifically, thanks to $\widehat{r}_l$ monotonically decreasing in $|\omega|$, i.e., $\widehat{r}_l(\omega_1) \geqslant \widehat{r}_l(\omega_2)$, for $\omega_1, \omega_2 \in \mathbb{R}^d$ with $|\omega_2| \geqslant |\omega_1|$, (4.71) can be established simply by showing that

$$\kappa_N(\omega) := |\omega|^2 \left| \frac{N-1}{N} \right|^{2\alpha} - \left|\omega - \nu\right|^2 \geqslant 0, \tag{4.72}$$

for all $\omega \in H \backslash B_\delta(0)$ and all $N \geqslant 2$. We first note that for $\omega \in H \backslash B_\delta(0)$ with $|\omega| > N^\alpha\delta$, (4.71) is trivially satisfied as the RHS of (4.71) equals zero (owing to $\left|\frac{\omega}{N^\alpha\delta}\right| > 1$ together with $\mathrm{supp}(\widehat{r}_l) \subseteq B_1(0)$). It hence suffices to prove (4.72) for $\omega \in H$ with $\delta \leqslant |\omega| \leqslant N^\alpha\delta$. To this end, fix $\tau \in [\delta, N^\alpha\delta]$, and define the spherical segment $\Xi_\tau := \{\omega \in H \mid |\omega| = \tau\}$. We then have

$$\kappa_N(\omega) = \tau^2 \left| \frac{N-1}{N} \right|^{2\alpha} - \left|\omega - \nu\right|^2 \geqslant \tau^2 \left| \frac{N-1}{N} \right|^{2\alpha} - \left|\omega^* - \nu\right|^2, \tag{4.73}$$

for $\omega \in \Xi_\tau$ and $N \geqslant 2$, where $\omega^* = (\tau, 0, \ldots, 0) \in \Xi_\tau$. The inequality in (4.73) holds thanks to the mapping $\omega \mapsto |\omega - \nu|^2$, $\omega \in \Xi_\tau$, attaining

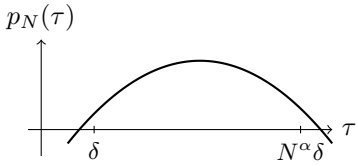

Fig. 4.10: The function $p_N(\tau)$ is quadratic in τ, with the coefficient of the highest-degree term negative. Establishing $p_N(\delta) \geqslant 0$ and $p_N(N^\alpha\delta) \geqslant 0$ therefore implies $p_N(\tau) \geqslant 0$, $\tau \in [\delta, N^\alpha\delta]$.

its maxima along the coordinate axes (see Fig. 4.9). Inserting

$$\begin{aligned}|\omega^* - \nu|^2 &= \left(\tau - \frac{\delta(1+2^{-1/2})}{d}\right)^2 + \frac{(d-1)\delta^2(1+2^{-1/2})^2}{d^2} \\ &= \tau^2 - \frac{\tau\delta(2+2^{1/2})}{d} + \frac{\delta^2(1+2^{-1/2})^2}{d}\end{aligned}$$

into (4.73) and rearranging terms yields

$$\kappa_N(\omega) \geqslant \underbrace{\tau^2\left(\left|\frac{N-1}{N}\right|^{2\alpha} - 1\right) + \frac{\tau\delta(2+2^{1/2})}{d} - \frac{\delta^2(1+2^{-1/2})^2}{d}}_{=:p_N(\tau)},$$

for all $\omega \in \Xi_\tau$ and all $N \geqslant 2$. This inequality shows that $\kappa_N(\omega)$ is lower-bounded—for $\omega \in \Xi_\tau$—by the 1-D function $p_N(\tau)$. Now, $p_N(\tau)$ is quadratic in τ, with the highest-degree coefficient $\left(\left|\frac{N-1}{N}\right|^{2\alpha} - 1\right)$ negative (owing to $\alpha = \log_2\left(\sqrt{d/(d-1/2)}\right) > 0$, for $d \geqslant 1$). Therefore, thanks to p_N, $N \geqslant 2$, being concave, establishing $p_N(\delta) \geqslant 0$ and $p_N(N^\alpha\delta) \geqslant 0$, for $N \geqslant 2$, implies $p_N(\tau) \geqslant 0$, for $\tau \in [\delta, N^\alpha\delta]$ and $N \geqslant 2$ (see Fig. 4.10), and thus (4.72), which completes the proof. It remains to show that $p_N(\delta) \geqslant 0$ and $p_N(N^\alpha\delta) \geqslant 0$, both for $N \geqslant 2$. We have

$$\begin{aligned}p_N(\delta) &= \delta^2\left(\left|\frac{N-1}{N}\right|^{2\alpha} - 1 + \frac{2+2^{1/2}}{d} - \frac{(1+2^{-1/2})^2}{d}\right) \\ &\geqslant \delta^2\left(2^{-2\alpha} - \frac{d-1/2}{d}\right) = 0,\end{aligned} \tag{4.74}$$

where the inequality in (4.74) is by $N \mapsto \left|\frac{N-1}{N}\right|^{2\alpha}$, for $N \geqslant 2$, monotonically increasing in N, and the equality is thanks to $\alpha = \log_2\left(\sqrt{d/(d-1/2)}\right)$, which is by assumption. Next, we have

$$\begin{aligned}\frac{p_N(N^\alpha\delta)}{\delta^2} &= \left|N-1\right|^{2\alpha} - N^{2\alpha} + \frac{N^\alpha(2+2^{1/2})}{d} - \frac{(1+2^{-1/2})^2}{d} \\ &\geqslant 1 - 2^{2\alpha} + \frac{2^\alpha(2+2^{1/2})}{d} - \frac{(1+2^{-1/2})^2}{d} \qquad (4.75) \\ &= 1 - \frac{d}{d-1/2} + \frac{\sqrt{d}(2+2^{1/2})}{d\sqrt{d-1/2}} - \frac{(1+2^{-1/2})^2}{d} \geqslant 0, \qquad (4.76)\end{aligned}$$

for all $d \geqslant 1$ and all $N \geqslant 2$, where (4.75) is by $N \mapsto (N-1)^{2\alpha} - N^{2\alpha} + d^{-1}N^\alpha(2+2^{1/2})$, for $N \geqslant 2$, monotonically increasing in N (owing to $\alpha = \log_2\left(\sqrt{d/(d-1/2)}\right) > 0$, for $d \geqslant 1$), and the equality in (4.76) is thanks to $\alpha = \log_2\left(\sqrt{d/(d-1/2)}\right)$. The inequality in (4.76) is established in Lemma 8 below. This completes the proof. □

It remains to show (4.76), which is accomplished through the following lemma.

Lemma 8. *For every $d \geqslant 1$ it holds that*

$$1 - \frac{d}{d-1/2} + \frac{\sqrt{d}(2+2^{1/2})}{d\sqrt{d-1/2}} - \frac{(1+2^{-1/2})^2}{d} \geqslant 0.$$

Proof. We start by multiplying the inequality by $d(d-1/2)$, which (after rearranging terms) yields

$$\sqrt{d(d-1/2)}\alpha \geqslant (d-1/2)\beta + d/2, \quad d \geqslant 1, \qquad (4.77)$$

where $\alpha := (2+2^{1/2})$ and $\beta := (1+2^{-1/2})^2$. Squaring (4.77) yields (again, after rearranging terms)

$$d^2 \underbrace{(\alpha^2 - \beta^2 - \beta - \frac{1}{4})}_{=0} + d\underbrace{(-\frac{\alpha^2}{2} + \beta^2 + \frac{\beta}{2})}_{\geqslant 4} - \underbrace{\frac{\beta^2}{4}}_{\leqslant 3} \geqslant 0, \quad d \geqslant 1,$$

which completes the proof. □

We proceed to sharpen the exponent $\alpha = \log_2(\sqrt{d/(d-1/2)})$ to $\alpha = 1$ for $d = 1$. The structure of the corresponding proof is similar to that of the proof for $d \geqslant 1$ with $\alpha = \log_2\big(\sqrt{d/(d-1/2)}\big)$. Specifically, we start by employing the arguments leading to (4.59) with N^α replaced by N. With this replacement $h_{n,N,\alpha,\delta}$ in (4.60) becomes $h_{n,N,\alpha,\delta}(\omega) := \sum_{\lambda_n \in \Lambda_n} |\widehat{g_{\lambda_n}}(\omega)|^2 \big(1 - \big|\widehat{r_l}\big(\frac{\omega - \nu_{\lambda_n}}{(N-1)\delta}\big)\big|^2\big)$, where, again, appropriate choice of the modulation factors $\{\nu_{\lambda_n}\}_{\lambda_n \in \Lambda_n} \subseteq \mathbb{R}$ will be key to establishing the inductive step. We start by defining Λ_n^+ to be the set of indices $\lambda_n \in \Lambda_n$ such that $\mathrm{supp}(\widehat{g_{\lambda_n}}) \subseteq [\delta, \infty)$, and take Λ_n^- to be the set of indices $\lambda_n \in \Lambda_n$ such that $\mathrm{supp}(\widehat{g_{\lambda_n}}) \subseteq (-\infty, -\delta]$ (see Assumption 1). Clearly, $\Lambda_n = \Lambda_n^+ \cup \Lambda_n^-$. Moreover, we define the modulation factors according to $\nu_{\lambda_n} := \delta$, for all $\lambda_n \in \Lambda_n^+$, and $\nu_{\lambda_n} := -\delta$, for all $\lambda_n \in \Lambda_n^-$. We then get

$$
\begin{aligned}
h_{n,N,\alpha,\delta}(\omega) &= \sum_{\lambda_n \in \Lambda_n} |\widehat{g_{\lambda_n}}(\omega)|^2 \Big(1 - \Big|\widehat{r_l}\Big(\frac{\omega - \nu_{\lambda_n}}{(N-1)\delta}\Big)\Big|^2\Big) \\
&= \sum_{\lambda_n \in \Lambda_n^+} |\widehat{g_{\lambda_n}}(\omega)|^2 \, \mathbb{1}_{[\delta,\,\infty)}(\omega) \Big(1 - \Big|\widehat{r_l}\Big(\frac{\omega - \delta}{(N-1)\delta}\Big)\Big|^2\Big) \qquad (4.78) \\
&+ \sum_{\lambda_n \in \Lambda_n^-} |\widehat{g_{\lambda_n}}(\omega)|^2 \, \mathbb{1}_{(-\infty,\,-\delta\,]}(\omega) \Big(1 - \Big|\widehat{r_l}\Big(\frac{\omega + \delta}{(N-1)\delta}\Big)\Big|^2\Big) \qquad (4.79) \\
&\leqslant \max\{1, B_n\} \, \mathbb{1}_{[\delta,\,\infty)}(\omega) \Big(1 - \Big|\widehat{r_l}\Big(\frac{\omega - \delta}{(N-1)\delta}\Big)\Big|^2\Big) \qquad (4.80) \\
&+ \max\{1, B_n\} \, \mathbb{1}_{(-\infty,\,-\delta\,]}(\omega) \Big(1 - \Big|\widehat{r_l}\Big(\frac{\omega + \delta}{(N-1)\delta}\Big)\Big|^2\Big), \qquad (4.81)
\end{aligned}
$$

where (4.78) and (4.79) are thanks to Assumption 1, and for the last step we employed (4.16). For the set $[\delta, \infty)$, we show in Lemma 9 below that

$$
\Big|\widehat{r_l}\Big(\frac{\omega - \delta}{(N-1)\delta}\Big)\Big| \geqslant \Big|\widehat{r_l}\Big(\frac{\omega}{N\delta}\Big)\Big|, \qquad \forall\, \omega \in [\delta, \infty),\ \forall N \geqslant 2. \qquad (4.82)
$$

This will allow us to deduce

$$
\Big|\widehat{r_l}\Big(\frac{\omega + \delta}{(N-1)\delta}\Big)\Big| \geqslant \Big|\widehat{r_l}\Big(\frac{\omega}{N\delta}\Big)\Big|, \qquad \forall\, \omega \in (-\infty, -\delta],\ \forall N \geqslant 2, \quad (4.83)
$$

simply by noting that

$$\begin{aligned}\Big|\widehat{r}_l\Big(\frac{\omega+\delta}{(N-1)\delta}\Big)\Big| &= \Big(1-\Big|\frac{\omega+\delta}{(N-1)\delta}\Big|\Big)_+^l = \Big(1-\Big|\frac{-(-\omega-\delta)}{(N-1)\delta}\Big|\Big)_+^l \\ &= \Big|\widehat{r}_l\Big(\frac{-\omega-\delta}{(N-1)\delta}\Big)\Big| \geqslant \Big|\widehat{r}_l\Big(\frac{-\omega}{N\delta}\Big)\Big| \qquad (4.84)\\ &= \Big(1-\Big|\frac{-\omega}{N\delta}\Big|\Big)_+^l = \Big|\widehat{r}_l\Big(\frac{\omega}{N\delta}\Big)\Big|,\end{aligned}$$

for $\omega \in (-\infty, -\delta]$. Here, the inequality in (4.84) is due to (4.82). Insertion of (4.82) into (4.80) and of (4.83) into (4.81) then yields

$$\begin{aligned}h_{n,N,\alpha,\delta}(\omega) &\leqslant \max\{1, B_n\}\, \mathbb{1}_{(-\infty,-\delta\,]\,\cup\,[\delta,\,\infty)}(\omega)\Big(1-\Big|\widehat{r}_l\Big(\frac{\omega}{N\delta}\Big)\Big|^2\Big)\\ &\leqslant \max\{1, B_n\}\Big(1-\Big|\widehat{r}_l\Big(\frac{\omega}{N\delta}\Big)\Big|^2\Big),\end{aligned}$$

for $\omega \in \mathbb{R}$, where the last inequality is thanks to $0 \leqslant \widehat{r}_l(\omega) \leqslant 1$, for $\omega \in \mathbb{R}$. This establishes (4.18)—in the 1-D case—for $\alpha = 1$ and completes the proof of statement i) in Theorem 3.

It remains to prove (4.82), which is done through the following lemma.

Lemma 9. *Let $\widehat{r}_l : \mathbb{R} \to \mathbb{R}$, $\widehat{r}_l(\omega) := (1-|\omega|)_+^l$, with $l > 1$. Then,*

$$\Big|\widehat{r}_l\Big(\frac{\omega-\delta}{(N-1)\delta}\Big)\Big| \geqslant \Big|\widehat{r}_l\Big(\frac{\omega}{N\delta}\Big)\Big|, \qquad \forall\, \omega \in [\delta, \infty),\ \forall N \geqslant 2. \quad (4.85)$$

Proof. We first note that for $\omega > N\delta$, (4.85) is trivially satisfied as the RHS of (4.85) equals zero (owing to $\big|\frac{\omega}{N\delta}\big| > 1$ together with $\mathrm{supp}(\widehat{r}_l) \subseteq B_1(0)$). It hence suffices to prove (4.85) for $\delta \leqslant \omega \leqslant N\delta$. The key idea of the proof is to employ a monotonicity argument. Specifically, thanks to $\widehat{r}_l$ monotonically decreasing in $|\omega|$, i.e., $\widehat{r}_l(\omega_1) \geqslant \widehat{r}_l(\omega_2)$, for $\omega_1, \omega_2 \in \mathbb{R}$ with $|\omega_2| \geqslant |\omega_1|$, (4.85) can be established simply by showing that

$$\Big|\frac{\omega-\delta}{(N-1)\delta}\Big| \leqslant \Big|\frac{\omega}{N\delta}\Big|, \qquad \forall\, \omega \in [\delta, N\delta\,],\ \forall N \geqslant 2,$$

which, by $\omega \in [\delta, N\delta]$, is equivalent to

$$\frac{\omega - \delta}{(N-1)\delta} \leqslant \frac{\omega}{N\delta}, \qquad \forall\, \omega \in [\delta, N\delta\,],\ \forall N \geqslant 2. \tag{4.86}$$

Rearranging terms in (4.86), we get

$$\omega \leqslant N\delta, \qquad \forall\, \omega \in [\delta, N\delta\,],\ \forall N \geqslant 2,$$

which completes the proof. □

Remark 7. *What makes the improved exponent α possible in the 1-D case is the absence of rotated orthants. Specifically, for $d = 1$, the filters $\{g_{\lambda_n}\}_{\lambda_n \in \Lambda_n}$ satisfy either supp$(\widehat{g_{\lambda_n}}) \subseteq (-\infty, -\delta]$ or supp$(\widehat{g_{\lambda_n}}) \subseteq [\delta, \infty)$, i.e., the support sets supp$(\widehat{g_{\lambda_n}})$ are located in one of the two half-spaces.*

4.7.3. Proof of statement ii) in Theorem 3

We need to show that there exist constants $C_{1,s}, C_{2,s} > 0$ (that are independent of N) such that

$$W_N(f) \leqslant C_{1,s} B_\Omega^N N^{-2s\alpha}, \quad \forall s \in (0, 1/2),\ \forall N \geqslant 1, \tag{4.87}$$

and

$$W_N(f) \leqslant C_{2,s} B_\Omega^N N^{-\alpha}, \quad \forall s \in [1/2, \infty),\ \forall N \geqslant 1. \tag{4.88}$$

Let us start by noting that

$$\max\{0, 1 - 2l|\omega|\} \leqslant (1 - |\omega|)_+^{2l}, \quad \omega \in \mathbb{R}^d, \tag{4.89}$$

where $l > \lfloor d/2 \rfloor + 1$, see Fig. 4.11. This implies

$$\begin{aligned} 1 - \left|\widehat{r}_l\Big(\frac{\omega}{N^\alpha \delta}\Big)\right|^2 &= 1 - \Big(1 - \Big|\frac{\omega}{N^\alpha \delta}\Big|\Big)_+^{2l} \\ &\leqslant 1 - \max\left\{0, 1 - \frac{2l\,|\omega|}{N^\alpha \delta}\right\} \\ &= 1 + \min\left\{0, \frac{2l\,|\omega|}{N^\alpha \delta} - 1\right\} \\ &= \min\left\{1, \frac{2l\,|\omega|}{N^\alpha \delta}\right\}, \quad \forall \omega \in \mathbb{R}^d. \end{aligned} \tag{4.90}$$

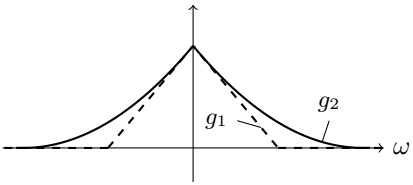

Fig. 4.11: Illustration of (4.89) in dimension $d = 1$. The functions $g_1(\omega) := \max\{0, 1 - 2l|\omega|\}$ (dashed line) and $g_2(\omega) := (1 - |\omega|)_+^{2l}$ (solid line) satisfy $g_1(\omega) \leqslant g_2(\omega)$, for $\omega \in \mathbb{R}$. Note that $l > \lfloor d/2 \rfloor + 1$.

The key idea of the proof of (4.87) is to upper-bound the integral on the RHS of (4.18) according to

$$\begin{aligned}
&\int_{\mathbb{R}^d} |\widehat{f}(\omega)|^2 \Big(1 - \Big|\widehat{r_l}\Big(\frac{\omega}{N^\alpha \delta}\Big)\Big|^2\Big) \mathrm{d}\omega \\
&\leqslant \int_{\mathbb{R}^d} |\widehat{f}(\omega)|^2 \min\left\{1, \frac{2l\,|\omega|}{N^\alpha \delta}\right\} \mathrm{d}\omega \qquad (4.91) \\
&= \int_{B_\tau(0)} |\widehat{f}(\omega)|^2 \frac{2l\,|\omega|}{N^\alpha \delta} \mathrm{d}\omega + \int_{\mathbb{R}^d \setminus B_\tau(0)} |\widehat{f}(\omega)|^2 \mathrm{d}\omega, \qquad (4.92)
\end{aligned}$$

where $\tau := \frac{N^\alpha \delta}{2l}$. Here, the inequality in (4.91) follows from (4.90), and (4.92) is owing to

$$\min\left\{1, \frac{2l\,|\omega|}{N^\alpha \delta}\right\} = \begin{cases} \frac{2l\,|\omega|}{N^\alpha \delta}, & |\omega| \leqslant \tau, \\ 1, & |\omega| > \tau. \end{cases}$$

Now, the first integral in (4.92) satisfies

$$\begin{aligned}
&\int_{B_\tau(0)} |\widehat{f}(\omega)|^2 \frac{2l\,|\omega|}{N^\alpha \delta} \mathrm{d}\omega = \frac{2l}{N^\alpha \delta} \int_{B_\tau(0)} |\widehat{f}(\omega)|^2 |\omega|^{1-2s} |\omega|^{2s} \mathrm{d}\omega \\
&\leqslant \frac{2l\,\tau^{1-2s}}{N^\alpha \delta} \int_{B_\tau(0)} |\widehat{f}(\omega)|^2 |\omega|^{2s} \mathrm{d}\omega \qquad (4.93) \\
&\leqslant \frac{2l\,\tau^{1-2s}}{N^\alpha \delta} \int_{B_\tau(0)} |\widehat{f}(\omega)|^2 (1 + |\omega|^2)^s \mathrm{d}\omega
\end{aligned}$$

$$\leqslant \Big(\frac{2l}{N^\alpha \delta}\Big)^{2s} \int_{B_\tau(0)} |\widehat{f}(\omega)|^2 (1+|\omega|^2)^s \,\mathrm{d}\omega, \tag{4.94}$$

where (4.93) is owing to $|\omega| \mapsto |\omega|^{1-2s}$ monotonically increasing in $|\omega|$ for $s \in (0, 1/2)$. For the second integral in (4.92), we have

$$\begin{aligned}
&\int_{\mathbb{R}^d \setminus B_\tau(0)} |\widehat{f}(\omega)|^2 \mathrm{d}\omega = \int_{\mathbb{R}^d \setminus B_\tau(0)} |\widehat{f}(\omega)|^2 |\omega|^{-2s} |\omega|^{2s} \mathrm{d}\omega \\
&\leqslant \tau^{-2s} \int_{\mathbb{R}^d \setminus B_\tau(0)} |\widehat{f}(\omega)|^2 \underbrace{|\omega|^{2s}}_{\leqslant (1+|\omega|^2)^s} \mathrm{d}\omega \qquad (4.95) \\
&\leqslant \tau^{-2s} \int_{\mathbb{R}^d \setminus B_\tau(0)} |\widehat{f}(\omega)|^2 (1+|\omega|^2)^s \mathrm{d}\omega \\
&\leqslant \Big(\frac{2l}{N^\alpha \delta}\Big)^{2s} \int_{\mathbb{R}^d \setminus B_\tau(0)} |\widehat{f}(\omega)|^2 (1+|\omega|^2)^s \,\mathrm{d}\omega, \qquad (4.96)
\end{aligned}$$

where (4.95) is thanks to

$$|\omega| \mapsto |\omega|^{-2s}, \quad \omega \in \mathbb{R}^d,$$

monotonically decreasing in $|\omega|$ for $s \in (0, 1/2)$. Inserting (4.94) and (4.96) into (4.92) establishes (4.87) with

$$C_{1,s} := (2l)^{2s} \delta^{-2s} \|f\|_{H^s}^2.$$

Next, we show (4.88) by noting that

$$\begin{aligned}
&\int_{\mathbb{R}^d} |\widehat{f}(\omega)|^2 \Big(1 - \Big|\widehat{r_l}\Big(\frac{\omega}{N^\alpha \delta}\Big)\Big|^2\Big) \mathrm{d}\omega \\
&\leqslant \int_{\mathbb{R}^d} |\widehat{f}(\omega)|^2 \min\Big\{1, \frac{2l\,|\omega|}{N^\alpha \delta}\Big\} \mathrm{d}\omega \qquad (4.97) \\
&\leqslant \frac{2l}{N^\alpha \delta} \int_{\mathbb{R}^d} |\widehat{f}(\omega)|^2 |\omega| \,\mathrm{d}\omega \\
&\leqslant \frac{2l}{N^\alpha \delta} \int_{\mathbb{R}^d} |\widehat{f}(\omega)|^2 (1+|\omega|^2)^s \,\mathrm{d}\omega = \frac{2l}{N^\alpha \delta} \|f\|_{H^s}^2,
\end{aligned}$$

where (4.97) is by (4.90), and the last inequality follows from $|\omega| \leqslant (1+|\omega|^2)^s$, for $\omega \in \mathbb{R}^d$ and $s \in [1/2, \infty)$. This establishes (4.88) with

$$C_{2,s} := (2l)\, \delta^{-1} \|f\|_{H^s}^2$$

and thereby completes the proof.

4.7.4. Proof of Proposition 8

Proposition 8. *Let Ω be the module-sequence* (4.1)*. Then,*

$$A_\Omega^N \|f\|_2^2 \leqslant \sum_{n=0}^{N-1} |||\Phi_\Omega^n(f)|||^2 + W_N(f) \leqslant B_\Omega^N \|f\|_2^2, \qquad (4.98)$$

for all $f \in L^2(\mathbb{R}^d)$ and all $N \geqslant 1$, where

$$A_\Omega^N = \prod_{k=1}^{N} \min\{1, A_k\}, \quad B_\Omega^N = \prod_{k=1}^{N} \max\{1, B_k\}.$$

Proof. We proceed by induction over N and start with the base case $N = 1$ which follows directly from the frame property (4.2) according to

$$A_\Omega^1 \|f\|_2^2 = \min\{1, A_1\}\|f\|_2^2 \leqslant A_1\|f\|_2^2 \leqslant \underbrace{\|f * \chi_0\|_2^2 + \sum_{\lambda_1 \in \Lambda_1} \|f * g_{\lambda_1}\|_2^2}_{= |||\Phi_\Omega^0(f)|||^2 + W_1(f)}$$
$$\leqslant B_1\|f\|_2^2 \leqslant \max\{1, B_1\}\|f\|_2^2 = B_\Omega^1 \|f\|_2^2, \qquad \forall\, f \in L^2(\mathbb{R}^d).$$

The inductive step is obtained as follows. Let $N > 1$ and suppose that (4.98) holds for $N - 1$, i.e.,

$$A_\Omega^{N-1} \|f\|_2^2 \leqslant \sum_{n=0}^{N-2} |||\Phi_\Omega^n(f)|||^2 + W_{N-1}(f) \leqslant B_\Omega^{N-1} \|f\|_2^2, \qquad (4.99)$$

for all $f \in L^2(\mathbb{R}^d)$. We start by noting that

$$\sum_{n=0}^{N-1} |||\Phi_\Omega^n(f)|||^2 + W_N(f) = \sum_{n=0}^{N-2} |||\Phi_\Omega^n(f)|||^2$$
$$+ \sum_{q \in \Lambda^{N-1}} \|(U[q]f) * \chi_{N-1}\|_2^2 + \sum_{q \in \Lambda^N} \|U[q]f\|_2^2, \qquad (4.100)$$

and proceed by examining the third term on the RHS of (4.100). Every path

$$\tilde{q} \in \Lambda^N = \underbrace{\Lambda_1 \times \ldots \times \Lambda_{N-1}}_{=\Lambda^{N-1}} \times \Lambda_N$$

of length N can be decomposed into a path $q \in \Lambda^{N-1}$ of length $N-1$ and an index $\lambda_N \in \Lambda_N$ according to $\tilde{q} = (q, \lambda_N)$. Thanks to (4.3) we have $U[\tilde{q}] = U[(q, \lambda_N)] = U_N[\lambda_N]U[q]$, which yields

$$\sum_{q \in \Lambda^N} \|U[q]f\|_2^2 = \sum_{q \in \Lambda^{N-1}} \sum_{\lambda_N \in \Lambda_N} \|(U[q]f) * g_{\lambda_N}\|_2^2. \tag{4.101}$$

Substituting the third term on the RHS of (4.100) by (4.101) and rearranging terms, we obtain

$$\begin{aligned}\sum_{n=0}^{N-1} |||\Phi_\Omega^n(f)|||^2 + W_N(f) &= \sum_{n=0}^{N-2} |||\Phi_\Omega^n(f)|||^2 \\ &+ \sum_{q \in \Lambda^{N-1}} \underbrace{\Big(\|(U[q]f) * \chi_{N-1}\|_2^2 + \sum_{\lambda_N \in \Lambda_N} \|(U[q]f) * g_{\lambda_N}\|_2^2 \Big)}_{=:\rho_N(U[q]f)}.\end{aligned}$$

Thanks to the frame property (4.2) and $U[q]f \in L^2(\mathbb{R}^d)$, which is by (3.11), we have

$$A_N \|U[q]f\|_2^2 \leqslant \rho_N(U[q]f) \leqslant B_N \|U[q]f\|_2^2,$$

and thus

$$\begin{aligned}&\min\{1, A_N\} \Big(\sum_{n=0}^{N-2} |||\Phi_\Omega^n(f)|||^2 + W_{N-1}(f) \Big) \qquad (4.102) \\ &\leqslant \sum_{n=0}^{N-1} |||\Phi_\Omega^n(f)|||^2 + W_N(f) \\ &\leqslant \max\{1, B_N\} \Big(\sum_{n=0}^{N-2} |||\Phi_\Omega^n(f)|||^2 + W_{N-1}(f) \Big), \qquad (4.103)\end{aligned}$$

where we employed the identity $\sum_{q\in\Lambda^{N-1}}\|U[q]f\|_2^2 = W_{N-1}(f)$. Invoking the induction hypothesis (4.99) in (4.102) and (4.103) and noting that

$$A_\Omega^N = \min\{1, A_N\}A_\Omega^{N-1}, \quad B_\Omega^N = \max\{1, B_N\}B_\Omega^{N-1},$$

completes the proof. □

4.7.5. Proof of Proposition 9

Proposition 9. *Let $\widehat{r_l} : \mathbb{R}^d \to \mathbb{R}$, $\widehat{r_l}(\omega) := (1-|\omega|)_+^l$, with $l > \lfloor d/2\rfloor + 1$, and α as defined in* (4.17)*. Then, we have*

$$\lim_{N\to\infty}\int_{\mathbb{R}^d}\left|\widehat{f}(\omega)\right|^2\Big(1-\Big|\widehat{r_l}\Big(\frac{\omega}{N^\alpha\delta}\Big)\Big|^2\Big)\mathrm{d}\omega = 0, \qquad (4.104)$$

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

Proof. We start by setting

$$d_{N,\alpha,\delta}(\omega) := \Big(1-\Big|\widehat{r_l}\Big(\frac{\omega}{N^\alpha\delta}\Big)\Big|^2\Big), \quad \omega\in\mathbb{R}^d,\, N\in\mathbb{N}.$$

Let $f\in L^2(\mathbb{R}^d)$. For every $\varepsilon>0$ there exists $R>0$ such that

$$\int_{\mathbb{R}^d\setminus\overline{B_R(0)}}|\widehat{f}(\omega)|^2\mathrm{d}\omega \leqslant \varepsilon/2,$$

where $\overline{B_R(0)}$ denotes the closed ball of radius R centered at the origin. Next, we employ Dini's Theorem (DiBenedetto, 2002, Theorem 7.3) to show that $(d_{N,\alpha,\delta})_{N\in\mathbb{N}}$ converges to the zero function $z_0(\omega) := 0$, $\omega\in\mathbb{R}^d$, uniformly on $\overline{B_R(0)}$. To this end, we note that (i) $d_{N,\alpha,\delta}$ is continuous as a composition of continuous functions, (ii) $z_0(\omega)=0$, for $\omega\in\mathbb{R}^d$, is, clearly, continuous, (iii) $d_{N,\alpha,\delta}(\omega)\geqslant d_{N+1,\alpha,\delta}(\omega)$, for $\omega\in\mathbb{R}^d$ and $N\in\mathbb{N}$, and (iv) $d_{N,\alpha,\delta}$ converges to z_0 pointwise on $\overline{B_R(0)}$, i.e., $\lim_{N\to\infty} d_{N,\alpha,\delta}(\omega) = z_0(\omega) = 0$, for $\omega\in\mathbb{R}^d$. This allows us

to conclude that there exists $N_0 \in \mathbb{N}$ (that depends on ε) such that $d_{N,\alpha,\delta}(\omega) \leqslant \frac{\varepsilon}{2\|f\|_2^2}$, for $\omega \in \overline{B_R(0)}$ and $N \geqslant N_0$, and we therefore get

$$\int_{\mathbb{R}^d} \left|\widehat{f}(\omega)\right|^2 d_{N,\alpha,\delta}(\omega)\mathrm{d}\omega = \int_{\mathbb{R}^d\setminus\overline{B_R(0)}} \left|\widehat{f}(\omega)\right|^2 \underbrace{d_{N,\alpha,\delta}(\omega)}_{\leqslant 1}\,\mathrm{d}\omega$$

$$+ \int_{\overline{B_R(0)}} \left|\widehat{f}(\omega)\right|^2 \underbrace{d_{N,\alpha,\delta}(\omega)}_{\leqslant \frac{\varepsilon}{2\|f\|_2^2}}\,\mathrm{d}\omega \leqslant \frac{\varepsilon}{2} + \frac{\varepsilon}{2\|f\|_2^2}\|\widehat{f}\|_2^2 = \varepsilon,$$

where in the last step we employed Parseval's formula. Since $\varepsilon > 0$ was arbitrary, we have (4.104), which completes the proof. □

4.7.6. Proof of Theorem 4

Wavelet case

We start by establishing (4.27) in statement i). The structure of the proof is similar to that of the proof of statement i) in Theorem 3 in Section 4.7.2, specifically we perform induction over N. Starting with the base case $N = 1$, we first note that $\mathrm{supp}(\widehat{\psi}) \subseteq [r^{-1}, r]$, $\widehat{g_j}(\omega) = \widehat{\psi}(r^{-j}\omega)$, for $j \geqslant 1$, and $\widehat{g_j}(\omega) = \widehat{\psi}(-r^{-|j|}\omega)$, for $j \leqslant -1$, all by assumption, imply

$$\mathrm{supp}(\widehat{g_j}) = \mathrm{supp}(\widehat{\psi}(r^{-j}\cdot)) \subseteq [r^{j-1}, r^{j+1}], \tag{4.105}$$

for $j \geqslant 1$, and

$$\mathrm{supp}(\widehat{g_j}) = \mathrm{supp}(\widehat{\psi}(-r^{-|j|}\cdot)) \subseteq [-r^{|j|+1}, -r^{|j|-1}], \tag{4.106}$$

for $j \leqslant -1$. We then get

$$W_1(f) = \sum_{j\in\mathbb{Z}\setminus\{0\}} \|f * g_j\|_2^2 = \int_{\mathbb{R}} \sum_{j\in\mathbb{Z}\setminus\{0\}} |\widehat{g_j}(\omega)|^2 |\widehat{f}(\omega)|^2 \mathrm{d}\omega \tag{4.107}$$

$$= \int_{\mathbb{R}} \sum_{j\geqslant 1} |\widehat{\psi}(r^{-j}\omega)|^2 |\widehat{f}(\omega)|^2 \mathrm{d}\omega + \int_{\mathbb{R}} \sum_{j\leqslant -1} |\widehat{\psi}(-r^{-|j|}\omega)|^2 |\widehat{f}(\omega)|^2 \mathrm{d}\omega$$

$$= \int_1^\infty \sum_{j\geqslant 1} |\widehat{\psi}(r^{-j}\omega)|^2 |\widehat{f}(\omega)|^2 \mathrm{d}\omega$$

$$+ \int_{-\infty}^{-1} \sum_{j\leqslant -1} |\widehat{\psi}(-r^{-|j|}\omega)|^2 |\widehat{f}(\omega)|^2 \mathrm{d}\omega \tag{4.108}$$

$$\leqslant \int_{\mathbb{R}\setminus[-1,1]} |\widehat{f}(\omega)|^2 \mathrm{d}\omega \leqslant \int_{\mathbb{R}} \big|\widehat{f}(\omega)\big|^2 (1 - |\widehat{r_l}(\omega)|^2) \mathrm{d}\omega, \tag{4.109}$$

where (4.107) is by Parseval's formula, and (4.108) is thanks to (4.105) and (4.106). The first inequality in (4.109) is owing to (4.24), and the second inequality is due to $\mathrm{supp}(\widehat{r_l}) \subseteq [-1,1]$ and $0 \leqslant \widehat{r_l}(\omega) \leqslant 1$, for $\omega \in \mathbb{R}$. The inductive step is obtained as follows. Let $N > 1$ and suppose that (4.27) holds for $N-1$, i.e.,

$$W_{N-1}(f) \leqslant \int_{\mathbb{R}} \big|\widehat{f}(\omega)\big|^2 \Big(1 - \Big|\widehat{r_l}\Big(\frac{\omega}{a^{N-2}}\Big)\Big|^2\Big) \mathrm{d}\omega, \tag{4.110}$$

for all $f \in L^2(\mathbb{R})$. We start by noting that every path $\tilde{q} \in (\mathbb{Z}\setminus\{0\})^N$ of length N can be decomposed into a path $q \in (\mathbb{Z}\setminus\{0\})^{N-1}$ of length $N-1$ and an index $j \in \mathbb{Z}\setminus\{0\}$ according to $\tilde{q} = (j,q)$. Thanks to (4.3) we have $U[\tilde{q}] = U[(j,q)] = U[q]U_1[j]$, which yields

$$\begin{aligned} W_N(f) &= \sum_{j\in\mathbb{Z}\setminus\{0\}} \sum_{q\,\in\,(\mathbb{Z}\setminus\{0\})^{N-1}} ||U[q](U_1[j]f)||_2^2 \\ &= \sum_{j\in\mathbb{Z}\setminus\{0\}} W_{N-1}(U_1[j]f). \end{aligned} \tag{4.111}$$

We proceed by examining the term $W_{N-1}(U_1[j]f)$ inside the sum in (4.111). Invoking the induction hypothesis (4.110) and employing Parseval's formula, we get

$$\begin{aligned} W_{N-1}(U_1[j]f) &\leqslant \int_{\mathbb{R}} \big|\widehat{U_1[j]f}(\omega)\big|^2 \Big(1 - \Big|\widehat{r_l}\Big(\frac{\omega}{a^{N-2}}\Big)\Big|^2\Big) \mathrm{d}\omega \\ &= \big(\|U_1[j]f\|_2^2 - \|(U_1[j]f) * r_{l,N-2}\|_2^2\big) \\ &= \big(\|f * g_j\|_2^2 - \||f * g_j| * r_{l,N-2}\|_2^2\big), \end{aligned} \tag{4.112}$$

where $r_{l,N-2}$ is the inverse Fourier transform of $\widehat{r_l}\big(\frac{\omega}{a^{N-2}}\big)$. Next, we note that $\widehat{r_l}\big(\frac{\omega}{a^{N-2}}\big)$ is a positive definite radial basis function (Wendland, 2004, Theorem 6.20) and hence by (Wendland, 2004, Theorem

6.18) $r_{l,N-2}(x) \geqslant 0$, for $x \in \mathbb{R}$. Furthermore, it follows from Lemma 6 in Section 4.7.2 that

$$\| |f * g_j| * r_{l,N-2} \|_2^2 \geqslant \| f * g_j * (M_{\nu_j} r_{l,N-2}) \|_2^2, \tag{4.113}$$

for all $\{\nu_j\}_{j \in \mathbb{Z}\setminus\{0\}} \subseteq \mathbb{R}$. Choosing the modulation factors $\{\nu_j\}_{j \in \mathbb{Z}\setminus\{0\}} \subseteq \mathbb{R}$ appropriately (see (4.117) below) will be key to establishing the inductive step. Using (4.112) and (4.113) to upper-bound the term $W_{N-1}(U_1[j]f)$ inside the sum in (4.111) yields

$$\begin{aligned} W_N(f) &\leqslant \sum_{j \in \mathbb{Z}\setminus\{0\}} \Big(\|f * g_j\|_2^2 - \|f * g_j * (M_{\nu_j} r_{l,N-2})\|_2^2 \Big) \\ &= \int_{\mathbb{R}} |\widehat{f}(\omega)|^2 h_{l,N-2}(\omega) \mathrm{d}\omega, \end{aligned} \tag{4.114}$$

where

$$h_{l,N-2}(\omega) := \sum_{j \in \mathbb{Z}\setminus\{0\}} |\widehat{g_j}(\omega)|^2 \Big(1 - \Big| \widehat{r_l}\Big(\frac{\omega - \nu_j}{a^{N-2}} \Big) \Big|^2 \Big). \tag{4.115}$$

In (4.114) we employed Parseval's formula together with $\widehat{M_\omega f} = T_\omega \widehat{f}$, for $f \in L^2(\mathbb{R})$ and $\omega \in \mathbb{R}$. The key step is now to establish—for appropriately chosen $\{\nu_j\}_{j \in \mathbb{Z}\setminus\{0\}} \subseteq \mathbb{R}$—the upper bound

$$h_{l,N-2}(\omega) \leqslant \Big(1 - \Big| \widehat{r_l}\Big(\frac{\omega}{a^{N-1}} \Big) \Big|^2 \Big), \qquad \forall\, \omega \in \mathbb{R}, \tag{4.116}$$

which then yields (4.27) and thereby completes the proof. To this end, we set $\eta := \frac{2r}{r^2+1}$,

$$\nu_j := r^j \eta, \quad j \geqslant 1, \qquad \nu_j := -r^{|j|}\eta, \quad j \leqslant -1, \tag{4.117}$$

and note that it suffices to prove (4.116) for $\omega \geqslant 0$, as

$$\begin{aligned} h_{l,N-2}(-\omega) &= \sum_{j \in \mathbb{Z}\setminus\{0\}} |\widehat{g_j}(-\omega)|^2 \Big(1 - \Big| \widehat{r_l}\Big(\frac{-\omega - \nu_j}{a^{N-2}} \Big) \Big|^2 \Big) \\ &= \sum_{j \leqslant -1} |\widehat{g_j}(-\omega)|^2 \Big(1 - \Big| \widehat{r_l}\Big(\frac{\omega + \nu_j}{a^{N-2}} \Big) \Big|^2 \Big) \end{aligned} \tag{4.118}$$

$$= \sum_{j\geqslant 1} |\widehat{g_{-j}}(-\omega)|^2 \Big(1 - \Big|\widehat{r_l}\Big(\frac{\omega+\nu_{-j}}{a^{N-2}}\Big)\Big|^2\Big)$$

$$= \sum_{j\geqslant 1} |\widehat{g_j}(\omega)|^2 \Big(1 - \Big|\widehat{r_l}\Big(\frac{\omega-\nu_j}{a^{N-2}}\Big)\Big|^2\Big) \tag{4.119}$$

$$= h_{l,N-2}(\omega), \quad \forall \omega \geqslant 0. \tag{4.120}$$

Here, (4.118) is thanks to $\widehat{g_j}(-\omega) = 0$, for $j \geqslant 1$ and $\omega \geqslant 0$, which is by (4.105), and (4.120) is owing to $\widehat{g_j}(\omega) = 0$, for $j \leqslant -1$ and $\omega \geqslant 0$, which is by (4.106). Moreover, in (4.118) we used that $\widehat{r_l}$ satisfies $\widehat{r_l}(-\omega) = \widehat{r_l}(\omega)$, for $\omega \in \mathbb{R}$, and (4.119) is thanks to

$$\widehat{g_{-j}}(-\omega) = \widehat{\psi}(r^{-|-j|}\omega) = \widehat{\psi}(r^{-j}\omega) = \widehat{g_j}(\omega), \qquad \forall\, \omega \in \mathbb{R},\ \forall\, j \geqslant 1,$$

as well as $\nu_{-j} = -r^j\eta = -\nu_j$, for $j \geqslant 1$. Now, let $\omega \in [0,1]$, and note that

$$h_{l,N-2}(\omega) = \sum_{j\in\mathbb{Z}\setminus\{0\}} |\widehat{g_j}(\omega)|^2 \Big(1 - \Big|\widehat{r_l}\Big(\frac{\omega-\nu_j}{a^{N-2}}\Big)\Big|^2\Big) = 0 \tag{4.121}$$

$$\leqslant 1 - \Big|\widehat{r_l}\Big(\frac{\omega}{a^{N-1}}\Big)\Big|^2, \quad \forall N \geqslant 2, \tag{4.122}$$

where the second equality in (4.121) is simply a consequence of $\widehat{g_j}(\omega) = 0$, for $j \in \mathbb{Z}\setminus\{0\}$ and $\omega \in [0,1]$, which, in turn, is by (4.105) and (4.106). The inequality in (4.122) is thanks to $0 \leqslant \widehat{r_l}(\omega) \leqslant 1$, for $\omega \in \mathbb{R}$. Next, let $\omega \in [1, r]$. Then, we have

$$h_{l,N-2}(\omega) = |\widehat{g_1}(\omega)|^2 \Big(1 - \Big|\widehat{r_l}\Big(\frac{\omega-r\eta}{a^{N-2}}\Big)\Big|^2\Big) \tag{4.123}$$

$$\leqslant |\widehat{g_1}(\omega)|^2 \Big(1 - \Big|\widehat{r_l}\Big(\frac{\omega-r\eta}{a^{N-2}}\Big)\Big|^2\Big)$$

$$+ \underbrace{\big(1 - |\widehat{g_1}(\omega)|^2\big)}_{\geqslant 0} \underbrace{\Big(1 - \Big|\widehat{r_l}\Big(\frac{\omega-\eta}{a^{N-2}}\Big)\Big|^2\Big)}_{\geqslant 0} \tag{4.124}$$

$$= 1 - \Big|\widehat{r_l}\Big(\frac{\omega-\eta}{a^{N-2}}\Big)\Big|^2$$

$$+ |\widehat{g_1}(\omega)|^2 \Big(\Big|\widehat{r_l}\Big(\frac{\omega-\eta}{a^{N-2}}\Big)\Big|^2 - \Big|\widehat{r_l}\Big(\frac{\omega-r\eta}{a^{N-2}}\Big)\Big|^2\Big), \tag{4.125}$$

where (4.123) is thanks to $\widehat{g_j}(\omega) = 0$, for $j \in \mathbb{Z}\backslash\{0,1\}$ and $\omega \in [1,r]$, which, in turn, is by (4.105) and (4.106). Moreover, (4.124) is owing to $|\widehat{g_1}(\omega)|^2 \in [0,1]$, which, in turn, is by (4.24) and $0 \leqslant \widehat{r_l}(\omega) \leqslant 1$, for $\omega \in \mathbb{R}$. Next, fix $j \geqslant 2$ and let $\omega \in [r^{j-1}, r^j]$. Then, we have

$$\begin{aligned} h_{l,N-2}(\omega) &= |\widehat{g_j}(\omega)|^2 \Big(1 - \Big|\widehat{r_l}\Big(\frac{\omega - r^j\eta}{a^{N-2}}\Big)\Big|^2\Big) \\ &+ \underbrace{|\widehat{g_{j-1}}(\omega)|^2}_{=(1-|\widehat{g_j}(\omega)|^2-|\widehat{\phi}(\omega)|^2)} \underbrace{\Big(1 - \Big|\widehat{r_l}\Big(\frac{\omega - r^{j-1}\eta}{a^{N-2}}\Big)\Big|^2\Big)}_{\geqslant 0} \qquad (4.126) \\ &\leqslant 1 - \Big|\widehat{r_l}\Big(\frac{\omega - r^{j-1}\eta}{a^{N-2}}\Big)\Big|^2 \\ &+ |\widehat{g_j}(\omega)|^2 \Big(\Big|\widehat{r_l}\Big(\frac{\omega - r^{j-1}\eta}{a^{N-2}}\Big)\Big|^2 - \Big|\widehat{r_l}\Big(\frac{\omega - r^j\eta}{a^{N-2}}\Big)\Big|^2\Big), \qquad (4.127) \end{aligned}$$

where (4.126) is thanks to i) $\widehat{g_{j'}}(\omega) = 0$, for $j' \in \mathbb{Z}\backslash\{0, j, j-1\}$ and $\omega \in [r^{j-1}, r^j]$, which, in turn, is by (4.105) and (4.106), and ii)

$$|\widehat{\phi}(\omega)|^2 + |\widehat{g_{j-1}}(\omega)|^2 + |\widehat{g_j}(\omega)|^2 = 1, \quad \forall \omega \in [r^{j-1}, r^j], \qquad (4.128)$$

which is a consequence of the Littlewood-Paley condition (4.24) and of (4.105) and (4.106). It follows from (4.125) and (4.127) that for every $j \geqslant 1$, we have

$$\begin{aligned} h_{l,N-2}(\omega) &\leqslant 1 - \Big|\widehat{r_l}\Big(\frac{\omega - r^{j-1}\eta}{a^{N-2}}\Big)\Big|^2 \\ &\quad + |\widehat{g_j}(\omega)|^2 \underbrace{\Big(\Big|\widehat{r_l}\Big(\frac{\omega - r^{j-1}\eta}{a^{N-2}}\Big)\Big|^2 - \Big|\widehat{r_l}\Big(\frac{\omega - r^j\eta}{a^{N-2}}\Big)\Big|^2\Big)}_{=:s(\omega)}, \end{aligned}$$

for $\omega \in [r^{j-1}, r^j]$. Next, we divide the interval $[r^{j-1}, r^j]$ into two intervals, namely $I_L := [r^{j-1}, \frac{r+1}{r^2+1} r^j]$ and $I_R := [\frac{r+1}{r^2+1} r^j, r^j]$, and note that $s(\omega) \geqslant 0$, for $\omega \in I_L$, and $s(\omega) \leqslant 0$, for $\omega \in I_R$, as $\widehat{r_l}$ is monotonically decreasing in $|\omega|$ and $|\omega - r^j\eta| \geqslant |\omega - r^{j-1}\eta|$, for $\omega \in I_L$, and $|\omega - r^j\eta| \leqslant |\omega - r^{j-1}\eta|$, for $\omega \in I_R$, respectively (see Fig. 4.12). For $\omega \in I_L$, we therefore have

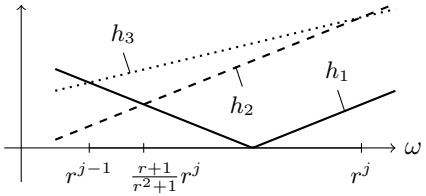

Fig. 4.12: The functions $h_1(\omega) := |\omega - r^j\eta|$ (solid line), $h_2(\omega) := |\omega - r^{j-1}\eta|$ (dashed line), and $h_3(\omega) := \frac{\omega}{a}$ (dotted line) satisfy $h_2 \leqslant h_1 \leqslant h_3$ on $I_L = [r^{j-1}, \frac{r+1}{r^2+1}r^j]$ and $h_1 \leqslant h_2 \leqslant h_3$ on $I_R = [\frac{r+1}{r^2+1}r^j, r^j]$.

$$\begin{aligned} h_{l,N-2}(\omega) &\leqslant 1 - \Big|\widehat{r}_l\Big(\frac{\omega - r^{j-1}\eta}{a^{N-2}}\Big)\Big|^2 + \underbrace{|\widehat{g_j}(\omega)|^2}_{\in\,[0,1]}\,\underbrace{s(\omega)}_{\geqslant 0} \\ &\leqslant 1 - \Big|\widehat{r}_l\Big(\frac{\omega - r^{j-1}\eta}{a^{N-2}}\Big)\Big|^2 + s(\omega) \\ &= 1 - \Big|\widehat{r}_l\Big(\frac{\omega - r^{j}\eta}{a^{N-2}}\Big)\Big|^2 \leqslant 1 - \Big|\widehat{r}_l\Big(\frac{\omega}{a^{N-1}}\Big)\Big|^2, \end{aligned}$$

where $|\widehat{g_j}(\omega)|^2 \in [0,1]$ follows from (4.128), and the last inequality is a consequence of $|\omega - r^j\eta| \leqslant \frac{\omega}{a}$, for $\omega \in I_L$, see Fig. 4.12. For $\omega \in I_R$, we have

$$\begin{aligned} h_{l,N-2}(\omega) &\leqslant 1 - \Big|\widehat{r}_l\Big(\frac{\omega - r^{j-1}\eta}{a^{N-2}}\Big)\Big|^2 + \underbrace{|\widehat{g_j}(\omega)|^2}_{\in\,[0,1]}\,\underbrace{s(\omega)}_{\leqslant 0} \\ &\leqslant 1 - \Big|\widehat{r}_l\Big(\frac{\omega - r^{j-1}\eta}{a^{N-2}}\Big)\Big|^2 \leqslant 1 - \Big|\widehat{r}_l\Big(\frac{\omega}{a^{N-1}}\Big)\Big|^2, \end{aligned}$$

where the last inequality now follows from $|\omega - r^{j-1}\eta| \leqslant \frac{\omega}{a}$, for $\omega \in I_R$, see Fig. 4.12. This completes the proof of (4.27).

Next, we establish (4.28). The proof is very similar to that of statement ii) in Theorem 3 in Section 4.7.3. We start by noting that (4.28) amounts to the existence of constants $C_{1,s}, C_{2,s} > 0$ (that are independent of N) such that

$$W_N(f) \leqslant C_{1,s}a^{-2sN}, \quad \forall s \in (0, 1/2),\ \forall N \geqslant 1, \tag{4.129}$$

and

$$W_N(f) \leqslant C_{2,s} a^{-N}, \quad \forall s \in [1/2, \infty),\ \forall N \geqslant 1, \tag{4.130}$$

where $a = \frac{r^2+1}{r^2-1}$, $r > 1$. The key idea of the proof of (4.129) is to upper-bound the integral on the RHS of (4.27) according to

$$\begin{aligned}
&\int_{\mathbb{R}} |\widehat{f}(\omega)|^2 \Big(1 - \Big|\widehat{r_l}\Big(\frac{\omega}{a^{N-1}}\Big)\Big|^2\Big) \mathrm{d}\omega \\
&\leqslant \int_{\mathbb{R}} |\widehat{f}(\omega)|^2 \min\left\{1, \frac{2l\,|\omega|}{a^{N-1}}\right\} \mathrm{d}\omega \qquad (4.131) \\
&= \int_{B_\tau(0)} |\widehat{f}(\omega)|^2 \frac{2l\,|\omega|}{a^{N-1}} \mathrm{d}\omega + \int_{\mathbb{R}\setminus B_\tau(0)} |\widehat{f}(\omega)|^2 \mathrm{d}\omega, \qquad (4.132)
\end{aligned}$$

where $\tau := \frac{a^{N-1}}{2l}$. Here, the inequality in (4.131) follows from (4.90), and (4.132) is owing to

$$\min\left\{1, \frac{2l\,|\omega|}{a^{N-1}}\right\} = \begin{cases} \frac{2l\,|\omega|}{a^{N-1}}, & |\omega| \leqslant \tau, \\ 1, & |\omega| > \tau. \end{cases}$$

Now, the first integral in (4.132) satisfies

$$\begin{aligned}
&\int_{B_\tau(0)} |\widehat{f}(\omega)|^2 \frac{2l\,|\omega|}{a^{N-1}} \mathrm{d}\omega = \frac{2l}{a^{N-1}} \int_{B_\tau(0)} |\widehat{f}(\omega)|^2 |\omega|^{1-2s} |\omega|^{2s} \mathrm{d}\omega \\
&\leqslant \frac{2l\,\tau^{1-2s}}{a^{N-1}} \int_{B_\tau(0)} |\widehat{f}(\omega)|^2 \underbrace{|\omega|^{2s}}_{\leqslant (1+|\omega|^2)^s} \mathrm{d}\omega \qquad (4.133) \\
&\leqslant \Big(\frac{2l}{a^{N-1}}\Big)^{2s} \int_{B_\tau(0)} |\widehat{f}(\omega)|^2 (1+|\omega|^2)^s \,\mathrm{d}\omega, \qquad (4.134)
\end{aligned}$$

where (4.133) is owing to $|\omega| \mapsto |\omega|^{1-2s}$ monotonically increasing in $|\omega|$ for $s \in (0, 1/2)$. For the second integral in (4.132), we have

$$\begin{aligned}
&\int_{\mathbb{R}\setminus B_\tau(0)} |\widehat{f}(\omega)|^2 \mathrm{d}\omega = \int_{\mathbb{R}\setminus B_\tau(0)} |\widehat{f}(\omega)|^2 |\omega|^{-2s} |\omega|^{2s} \mathrm{d}\omega \\
&\leqslant \tau^{-2s} \int_{\mathbb{R}\setminus B_\tau(0)} |\widehat{f}(\omega)|^2 \underbrace{|\omega|^{2s}}_{\leqslant (1+|\omega|^2)^s} \mathrm{d}\omega \qquad (4.135)
\end{aligned}$$

$$\leqslant \Big(\frac{2l}{a^{N-1}}\Big)^{2s} \int_{\mathbb{R}\setminus B_\tau(0)} |\widehat{f}(\omega)|^2 (1+|\omega|^2)^s \,\mathrm{d}\omega, \tag{4.136}$$

where (4.135) is thanks to $|\omega| \mapsto |\omega|^{-2s}$ monotonically decreasing in $|\omega|$ for $s \in (0, 1/2)$. Inserting (4.134) and (4.136) into (4.132) establishes (4.129) with

$$C_{1,s} := (2l)^{2s} a^{2s} \|f\|_{H^s}^2.$$

Next, we show (4.130) by noting that

$$\begin{aligned}
&\int_{\mathbb{R}} |\widehat{f}(\omega)|^2 \Big(1 - \Big|\widehat{r_l}\Big(\frac{\omega}{a^{N-1}}\Big)\Big|^2\Big) \mathrm{d}\omega \\
&\leqslant \int_{\mathbb{R}} |\widehat{f}(\omega)|^2 \min\Big\{1, \frac{2l\,|\omega|}{a^{N-1}}\Big\} \mathrm{d}\omega \qquad (4.137) \\
&\leqslant \frac{2l}{a^{N-1}} \int_{\mathbb{R}} |\widehat{f}(\omega)|^2 |\omega| \,\mathrm{d}\omega \\
&\leqslant \frac{2l}{a^{N-1}} \int_{\mathbb{R}} |\widehat{f}(\omega)|^2 (1+|\omega|^2)^s \,\mathrm{d}\omega = \frac{2l}{a^{N-1}} \|f\|_{H^s}^2,
\end{aligned}$$

where (4.137) is by (4.90), and the last inequality follows from

$$|\omega| \leqslant (1+|\omega|^2)^s, \quad \forall\, \omega \in \mathbb{R},\ \forall\, s \in [1/2, \infty).$$

This establishes (4.130) with

$$C_{2,s} := 2la \|f\|_{H^s}^2$$

and thereby completes the proof of statement i).

Weyl-Heisenberg case

We proceed to the proof of statement ii), again, effected by induction over N. Specifically, we first establish (4.30) by employing the same arguments as those leading to (4.114) with a^{N-2} (where a is defined in (4.26)) replaced by $a^{N-2}\delta$ (where a is defined in (4.29)). With this replacement $h_{l,N-2}$ in (4.115) becomes

$$h_{l,N-2}(\omega) := \sum_{k \in \mathbb{Z}\setminus\{0\}} |\widehat{g_k}(\omega)|^2 \Big(1 - \Big|\widehat{r_l}\Big(\frac{\omega - \nu_k}{a^{N-2}\delta}\Big)\Big|^2\Big), \tag{4.138}$$

where, again, appropriate choice of the modulation factors $\{\nu_k\}_{k\in\mathbb{Z}\setminus\{0\}} \subseteq \mathbb{R}$ (see (4.142) below) will be key in establishing the inductive step. Here, we note that the functions $\widehat{g_k}$ in (4.138) satisfy $\widehat{g_k}(\omega) = \widehat{g}(\omega - (Rk+\delta))$, for $k \geqslant 1$, $\widehat{g_k}(\omega) = \widehat{g}(\omega + (R|k|+\delta))$, for $k \leqslant -1$, by assumption, as well as

$$\begin{aligned}\operatorname{supp}(\widehat{g_k}) &= \operatorname{supp}(\widehat{g}(\cdot - (Rk+\delta))) \\ &\subseteq [\delta + R(k-1), \delta + R(k+1)], \quad k \geqslant 1, \end{aligned} \tag{4.139}$$

and

$$\begin{aligned}\operatorname{supp}(\widehat{g_k}) &= \operatorname{supp}(\widehat{g}(\cdot + (R|k|+\delta))) \\ &\subseteq [-(\delta + R(|k|+1)), -(\delta + R(|k|-1))], \end{aligned} \tag{4.140}$$

for $k \leqslant -1$, where (4.139) and (4.140) follow from $\operatorname{supp}(\widehat{g}) \subseteq [-R, R]$, which is by assumption. It remains to establish the equivalent of (4.116), namely

$$h_{l,N-2}(\omega) \leqslant 1 - \Big|\widehat{r_l}\Big(\frac{\omega}{a^{N-1}\delta}\Big)\Big|^2, \quad \forall \omega \in \mathbb{R}. \tag{4.141}$$

To this end, we set $\eta := \frac{R^2}{R+2\delta}$,

$$\nu_k := \delta + Rk - \eta, \quad \forall k \geqslant 1, \qquad \nu_k := -\nu_{|k|}, \quad \forall k \leqslant -1, \tag{4.142}$$

and note that it suffices to establish (4.141) for $\omega \geqslant 0$, thanks to

$$\begin{aligned} h_{l,N-2}(-\omega) &= \sum_{k\in\mathbb{Z}\setminus\{0\}} |\widehat{g_k}(-\omega)|^2 \Big(1 - \Big|\widehat{r_l}\Big(\frac{-\omega-\nu_k}{a^{N-2}\delta}\Big)\Big|^2\Big) \\ &= \sum_{k\leqslant -1} |\widehat{g_k}(-\omega)|^2 \Big(1 - \Big|\widehat{r_l}\Big(\frac{\omega+\nu_k}{a^{N-2}\delta}\Big)\Big|^2\Big) \qquad (4.143) \\ &= \sum_{k\geqslant 1} |\widehat{g_{-k}}(-\omega)|^2 \Big(1 - \Big|\widehat{r_l}\Big(\frac{\omega+\nu_{-k}}{a^{N-2}\delta}\Big)\Big|^2\Big) \\ &= \sum_{k\geqslant 1} |\widehat{g_k}(\omega)|^2 \Big(1 - \Big|\widehat{r_l}\Big(\frac{\omega-\nu_k}{a^{N-2}\delta}\Big)\Big|^2\Big) \qquad (4.144) \\ &= h_{l,N-2}(\omega), \quad \forall \omega \geqslant 0. \qquad (4.145) \end{aligned}$$

Here, (4.143) follows from $\widehat{g_k}(-\omega) = 0$, for $k \geqslant 1$ and $\omega \geqslant 0$, which, in turn, is by (4.139), and (4.145) is owing to $\widehat{g_k}(\omega) = 0$, for $k \leqslant -1$ and $\omega \geqslant 0$, which is by (4.140). Moreover, in (4.143) we used that $\widehat{r_l}$ satisfies $\widehat{r_l}(-\omega) = \widehat{r_l}(\omega)$, for $\omega \in \mathbb{R}$, and (4.144) is thanks to $\nu_{-k} = -\nu_k$, for $k \geqslant 1$, and

$$\begin{aligned}\widehat{g_{-k}}(-\omega) &= \widehat{g}(-\omega + (R|-k| + \delta)) = \widehat{g}(-(\omega - (Rk + \delta))) \\ &= \widehat{g}(\omega - (Rk + \delta)) = \widehat{g_k}(\omega), \quad \forall \omega \in \mathbb{R}, \forall\, k \geqslant 1,\end{aligned}$$

where we used $\widehat{g}(-\omega) = \widehat{g}(\omega)$, for $\omega \in \mathbb{R}$, which is by assumption. Now, let $\omega \in [0, \delta]$, and note that

$$h_{l,N-2}(\omega) = 0 \leqslant 1 - \Big|\widehat{r_l}\Big(\frac{\omega}{a^{N-1}\delta}\Big)\Big|^2, \quad \forall N \geqslant 2, \tag{4.146}$$

where the equality in (4.146) is a consequence of (4.139) and (4.140), and the inequality is thanks to $0 \leqslant \widehat{r_l}(\omega) \leqslant 1$, for $\omega \in \mathbb{R}$. Next, let $\omega \in [\delta, \delta + R]$. Then, we have

$$\begin{aligned}h_{l,N-2}(\omega) &= |\widehat{g_1}(\omega)|^2\Big(1 - \Big|\widehat{r_l}\Big(\frac{\omega - \nu_1}{a^{N-2}\delta}\Big)\Big|^2\Big) && (4.147)\\ &\leqslant |\widehat{g_1}(\omega)|^2\Big(1 - \Big|\widehat{r_l}\Big(\frac{\omega - \nu_1}{a^{N-2}\delta}\Big)\Big|^2\Big) \\ &\quad + \underbrace{(1 - |\widehat{g_1}(\omega)|^2)}_{\geqslant 0}\,\underbrace{\Big(1 - \Big|\widehat{r_l}\Big(\frac{\omega - (\delta - \nu)}{a^{N-2}\delta}\Big)\Big|^2\Big)}_{\geqslant 0} && (4.148)\\ &= 1 - \Big|\widehat{r_l}\Big(\frac{\omega - (\delta - \nu)}{a^{N-2}\delta}\Big)\Big|^2 \\ &\quad + |\widehat{g_1}(\omega)|^2\Big(\Big|\widehat{r_l}\Big(\frac{\omega - (\delta - \nu)}{a^{N-2}\delta}\Big)\Big|^2 - \Big|\widehat{r_l}\Big(\frac{\omega - \nu_1}{a^{N-2}\delta}\Big)\Big|^2\Big), && (4.149)\end{aligned}$$

where (4.147) is thanks to $\widehat{g_k}(\omega) = 0$, for $k \in \mathbb{Z}\backslash\{0, 1\}$ and $\omega \in [\delta, \delta + R]$, which, in turn, is by (4.139) and (4.140). Moreover, (4.148) is owing to $|\widehat{g_1}(\omega)|^2 \in [0, 1]$, which, in turn, is by (4.25), and $0 \leqslant \widehat{r_l}(\omega) \leqslant 1$, for $\omega \in \mathbb{R}$. Next, fix $k \geqslant 2$, and let $\omega \in [\delta + R(k-1), \delta + Rk]$.

Then, we have

$$h_{l,N-2}(\omega) = |\widehat{g_k}(\omega)|^2\Big(1 - \Big|\widehat{r_l}\Big(\frac{\omega-\nu_k}{a^{N-2}\delta}\Big)\Big|^2\Big)$$
$$+ \underbrace{|\widehat{g_{k-1}}(\omega)|^2}_{=(1-|\widehat{g_k}(\omega)|^2-|\widehat{\phi}(\omega)|^2)} \underbrace{\Big(1 - \Big|\widehat{r_l}\Big(\frac{\omega-\nu_{k-1}}{a^{N-2}\delta}\Big)\Big|^2\Big)}_{\geqslant 0} \tag{4.150}$$
$$\leqslant 1 - \Big|\widehat{r_l}\Big(\frac{\omega-\nu_{k-1}}{a^{N-2}\delta}\Big)\Big|^2$$
$$+ |\widehat{g_k}(\omega)|^2\Big(\Big|\widehat{r_l}\Big(\frac{\omega-\nu_{k-1}}{a^{N-2}\delta}\Big)\Big|^2 - \Big|\widehat{r_l}\Big(\frac{\omega-\nu_k}{a^{N-2}\delta}\Big)\Big|^2\Big), \tag{4.151}$$

where (4.150) is thanks to i) $\widehat{g_{k'}}(\omega) = 0$, for $k' \in \mathbb{Z}\backslash\{0, k, k-1\}$ and $\omega \in [\delta + R(k-1), \delta + Rk]$, which, in turn, is by (4.139) and (4.140), and ii)

$$|\widehat{\chi}(\omega)|^2 + |\widehat{g_{k-1}}(\omega)|^2 + |\widehat{g_k}(\omega)|^2 = 1, \tag{4.152}$$

for all $\omega \in [\delta + R(k-1), \delta + Rk]$, which is a consequence of the Littlewood-Paley condition (4.25) and of (4.139) and (4.140). It follows from (4.149) and (4.151) that for $k \geqslant 1$, we have

$$h_{l,N-2}(\omega) \leqslant 1 - \Big|\widehat{r_l}\Big(\frac{\omega-\nu_{k-1}}{a^{N-2}\delta}\Big)\Big|^2$$
$$+ |\widehat{g_k}(\omega)|^2 \underbrace{\Big(\Big|\widehat{r_l}\Big(\frac{\omega-\nu_{k-1}}{a^{N-2}\delta}\Big)\Big|^2 - \Big|\widehat{r_l}\Big(\frac{\omega-\nu_k}{a^{N-2}\delta}\Big)\Big|^2\Big)}_{=:s(\omega)},$$

for $\omega \in [\delta + R(k-1), \delta + Rk]$, where $\nu_0 := (\delta - \nu)$. Next, we divide the interval $[\delta + R(k-1), \delta + Rk]$ into two intervals, namely $I_L := [\delta + R(k-1), \tau]$ and $I_R := [\tau, \delta + Rk]$, where $\tau := \delta + Rk - R/2 - \eta$, and note that $s(\omega) \geqslant 0$, for $\omega \in I_L$, and $s(\omega) \leqslant 0$, for $\omega \in I_R$, as $\widehat{r_l}$ is monotonically decreasing in $|\omega|$ and $|\omega - \nu_k| \geqslant |\omega - \nu_{k-1}|$, for $\omega \in I_L$, and $|\omega - \nu_k| \leqslant |\omega - \nu_{k-1}|$, for $\omega \in I_R$, respectively (see Fig. 4.13). For $\omega \in I_L$, we therefore have

$$h_{l,N-2}(\omega) \leqslant 1 - \Big|\widehat{r_l}\Big(\frac{\omega-\nu_{k-1}}{a^{N-2}\delta}\Big)\Big|^2 + \underbrace{|\widehat{g_k}(\omega)|^2}_{\in[0,1]} \underbrace{s(\omega)}_{\geqslant 0}$$

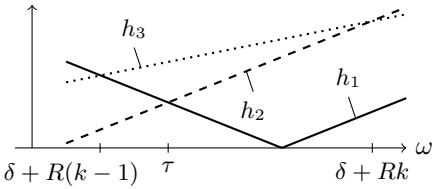

Fig. 4.13: The functions $h_1(\omega) := |\omega - \nu_k|$ (solid line), $h_2(\omega) := |\omega - \nu_{k-1}|$ (dashed line), and $h_3(\omega) = \frac{\omega}{a}$ (dotted line) satisfy $h_2 \leqslant h_1 \leqslant h_3$ on $I_L = [\delta + R(k-1), \tau]$ and $h_1 \leqslant h_2 \leqslant h_3$ on $I_R = [\tau, \delta + Rk]$.

$$\begin{aligned} &\leqslant 1 - \left|\widehat{r}_l\Big(\frac{\omega - \nu_{k-1}}{a^{N-2}\delta}\Big)\right|^2 + s(\omega) \\ &= 1 - \left|\widehat{r}_l\Big(\frac{\omega - \nu_k}{a^{N-2}\delta}\Big)\right|^2 \leqslant 1 - \left|\widehat{r}_l\Big(\frac{\omega}{a^{N-1}\delta}\Big)\right|^2, \end{aligned}$$

where $|\widehat{g_k}(\omega)|^2 \in [0,1]$ follows from (4.152), and the last inequality is by $|\omega - \nu_k| \leqslant \frac{\omega}{a}$, for $\omega \in I_L$ (see Fig. 4.13). For the interval $\omega \in I_R$, we have

$$\begin{aligned} h_{l,N-2}(\omega) &\leqslant 1 - \left|\widehat{r}_l\Big(\frac{\omega - \nu_{k-1}}{a^{N-2}\delta}\Big)\right|^2 + \underbrace{|\widehat{g_k}(\omega)|^2}_{\in [0,1]} \underbrace{s(\omega)}_{\leqslant 0} \\ &\leqslant 1 - \left|\widehat{r}_l\Big(\frac{\omega - \nu_{k-1}}{a^{N-2}\delta}\Big)\right|^2 \leqslant 1 - \left|\widehat{r}_l\Big(\frac{\omega}{a^{N-1}\delta}\Big)\right|^2, \end{aligned}$$

where the last inequality is by $|\omega - \nu_{k-1}| \leqslant \frac{\omega}{a}$, for $\omega \in I_R$ (see Fig. 4.13). This completes the proof of (4.30).

Next, we establish (4.31). The proof is very similar to that of statement ii) in Theorem 3 in Section 4.7.3. We start by noting that (4.31) amounts to the existence of constants $C_{1,s}, C_{2,s} > 0$ (that are independent of N) such that

$$W_N(f) \leqslant C_{1,s} a^{-2sN}, \quad \forall s \in (0, 1/2),\ \forall N \geqslant 1, \tag{4.153}$$

and

$$W_N(f) \leqslant C_{2,s} a^{-N}, \quad \forall s \in [1/2, \infty),\ \forall N \geqslant 1, \tag{4.154}$$

where $a = \frac{1}{2} + \frac{\delta}{R}$, $\delta \geqslant \frac{R}{2}$. The key idea of the proof of (4.153) is to upper-bound the integral on the RHS of (4.30) according to

$$\begin{aligned}
&\int_{\mathbb{R}} |\widehat{f}(\omega)|^2 \Big(1 - \Big|\widehat{r}_l\Big(\frac{\omega}{a^{N-1}R}\Big)\Big|^2\Big) \mathrm{d}\omega \\
&\leqslant \int_{\mathbb{R}} |\widehat{f}(\omega)|^2 \min\left\{1, \frac{2l\,|\omega|}{a^{N-1}R}\right\} \mathrm{d}\omega \qquad (4.155) \\
&= \int_{B_\tau(0)} |\widehat{f}(\omega)|^2 \frac{2l\,|\omega|}{a^{N-1}R} \mathrm{d}\omega + \int_{\mathbb{R}\setminus B_\tau(0)} |\widehat{f}(\omega)|^2 \mathrm{d}\omega, \qquad (4.156)
\end{aligned}$$

where

$$\tau := \frac{a^{N-1}R}{2l}.$$

Here, the inequality in (4.155) follows from (4.90), and (4.156) is owing to

$$\min\left\{1, \frac{2l\,|\omega|}{a^{N-1}R}\right\} = \begin{cases} \frac{2l\,|\omega|}{a^{N-1}R}, & |\omega| \leqslant \tau, \\ 1, & |\omega| > \tau. \end{cases}$$

Now, the first integral in (4.156) satisfies

$$\begin{aligned}
&\int_{B_\tau(0)} |\widehat{f}(\omega)|^2 \frac{2l\,|\omega|}{a^{N-1}R} \mathrm{d}\omega = \frac{2l}{a^{N-1}R} \int_{B_\tau(0)} |\widehat{f}(\omega)|^2 |\omega|^{1-2s} |\omega|^{2s} \mathrm{d}\omega \\
&\leqslant \frac{2l\,\tau^{1-2s}}{a^{N-1}R} \int_{B_\tau(0)} |\widehat{f}(\omega)|^2 \underbrace{|\omega|^{2s}}_{\leqslant (1+|\omega|^2)^s} \mathrm{d}\omega \qquad (4.157) \\
&\leqslant \Big(\frac{2l}{a^{N-1}R}\Big)^{2s} \int_{B_\tau(0)} |\widehat{f}(\omega)|^2 (1+|\omega|^2)^s \,\mathrm{d}\omega, \qquad (4.158)
\end{aligned}$$

where (4.157) is owing to $|\omega| \mapsto |\omega|^{1-2s}$ monotonically increasing in $|\omega|$ for $s \in (0, 1/2)$. For the second integral in (4.156), we have

$$\begin{aligned}
&\int_{\mathbb{R}\setminus B_\tau(0)} |\widehat{f}(\omega)|^2 \mathrm{d}\omega = \int_{\mathbb{R}\setminus B_\tau(0)} |\widehat{f}(\omega)|^2 |\omega|^{-2s} |\omega|^{2s} \mathrm{d}\omega \\
&\leqslant \tau^{-2s} \int_{\mathbb{R}\setminus B_\tau(0)} |\widehat{f}(\omega)|^2 \underbrace{|\omega|^{2s}}_{\leqslant (1+|\omega|^2)^s} \mathrm{d}\omega \qquad (4.159) \\
&\leqslant \Big(\frac{2l}{a^{N-1}R}\Big)^{2s} \int_{\mathbb{R}\setminus B_\tau(0)} |\widehat{f}(\omega)|^2 (1+|\omega|^2)^s \,\mathrm{d}\omega, \qquad (4.160)
\end{aligned}$$

where (4.159) is thanks to

$$|\omega| \mapsto |\omega|^{-2s}, \quad \omega \in \mathbb{R},$$

monotonically decreasing in $|\omega|$ for $s \in (0, 1/2)$. Inserting (4.158) and (4.160) into (4.156) establishes (4.153) with

$$C_{1,s} := (2l)^{2s} a^{2s} R^{-2s} \|f\|_{H^s}^2.$$

Next, we show (4.154) by noting that

$$\begin{aligned}
&\int_{\mathbb{R}} |\widehat{f}(\omega)|^2 \Big(1 - \Big|\widehat{r_l}\Big(\frac{\omega}{a^{N-1}R}\Big)\Big|^2\Big) \mathrm{d}\omega \\
&\leqslant \int_{\mathbb{R}} |\widehat{f}(\omega)|^2 \min\left\{1, \frac{2l\,|\omega|}{a^{N-1}R}\right\} \mathrm{d}\omega \qquad (4.161) \\
&\leqslant \frac{2l}{a^{N-1}R} \int_{\mathbb{R}} |\widehat{f}(\omega)|^2 |\omega| \,\mathrm{d}\omega \\
&\leqslant \frac{2l}{a^{N-1}R} \int_{\mathbb{R}} |\widehat{f}(\omega)|^2 (1 + |\omega|^2)^s \,\mathrm{d}\omega \\
&= \frac{2l}{a^{N-1}R} \|f\|_{H^s}^2,
\end{aligned}$$

where (4.161) is by (4.90), and the last inequality follows from $|\omega| \leqslant (1 + |\omega|^2)^s$, for $\omega \in \mathbb{R}$ and $s \in [1/2, \infty)$. This establishes (4.154) with

$$C_{2,s} := 2laR^{-1} \|f\|_{H^s}^2$$

and thereby completes the proof of statement ii).

4.7.7. Proof of Corollary 2

We start with statement i) and note that $A_\Omega^N = B_\Omega^N = 1$, $N \in \mathbb{N}$, by assumption. Let $f \in L^2(\mathbb{R}^d)$ with $\mathrm{supp}(\widehat{f}) \subseteq B_L(0)$. Then, by Proposition 8 in Section 4.7.4 together with $\lim_{N\to\infty} W_N(f) = 0$, for $f \in L^2(\mathbb{R}^d)$, which follows from Proposition 9 in Section 4.7.5, we

have

$$\begin{aligned}\|f\|_2^2 = |||\Phi_\Omega(f)|||^2 &= \sum_{n=0}^{\infty} |||\Phi_\Omega^n(f)|||^2 \\ &\geqslant \sum_{n=0}^{N} |||\Phi_\Omega^n(f)|||^2 = \|f\|_2^2 - W_{N+1}(f) \qquad (4.162) \\ &\geqslant \int_{\mathbb{R}^d} |\widehat{f}(\omega)|^2 \Big|\widehat{r_l}\Big(\frac{\omega}{(N+1)^\alpha \delta}\Big)\Big|^2 \mathrm{d}\omega \qquad (4.163) \\ &= \int_{B_L(0)} |\widehat{f}(\omega)|^2 \Big|\widehat{r_l}\Big(\frac{\omega}{(N+1)^\alpha \delta}\Big)\Big|^2 \mathrm{d}\omega, \qquad (4.164)\end{aligned}$$

where (4.162) is by the lower bound in (4.98), (4.163) is thanks to Parseval's formula and (4.18), and (4.164) follows from f being L-band-limited. Next, thanks to $\widehat{r_l}$ monotonically decreasing in $|\omega|$, we get

$$\Big|\widehat{r_l}\Big(\frac{\omega}{(N+1)^\alpha \delta}\Big)\Big|^2 \geqslant \Big|\widehat{r_l}\Big(\frac{L}{(N+1)^\alpha \delta}\Big)\Big|^2, \qquad \forall\, \omega \in B_L(0). \quad (4.165)$$

Employing (4.165) in (4.164), we obtain

$$\begin{aligned}\|f\|_2^2 &\geqslant \Big|\widehat{r_l}\Big(\frac{L}{(N+1)^\alpha \delta}\Big)\Big|^2 \|f\|_2^2 = \Big(1 - \frac{L}{(N+1)^\alpha \delta}\Big)_+^{2l} \|f\|_2^2 \qquad (4.166) \\ &= \Big(1 - \frac{L}{(N+1)^\alpha \delta}\Big)^{2l} \|f\|_2^2 \geqslant (1-\varepsilon)\|f\|_2^2, \qquad (4.167)\end{aligned}$$

where in (4.166) we used Parseval's formula, the equality in (4.167) is due to $L \leqslant (N+1)^\alpha \delta$, which, in turn, is by (4.33), and the inequality in (4.167) is also by (4.33) (upon rearranging terms). This establishes (4.32) and thereby completes the proof.

The proof of statement ii) is very similar to that of statement i). Again, we start by noting that $A_\Omega^N = B_\Omega^N = 1$, $N \in \mathbb{N}$, by assumption. Let $f \in L^2(\mathbb{R})$ with $\mathrm{supp}(\widehat{f}) \subseteq B_L(0)$. Then, by Proposition 8 in Section 4.7.4 together with $\lim_{N\to\infty} W_N(f) = 0$, for $f \in L^2(\mathbb{R})$, we have

$$\|f\|_2^2 = |||\Phi_\Omega(f)|||^2 = \sum_{n=0}^{\infty} |||\Phi_\Omega^n(f)|||^2$$

$$\geqslant \sum_{n=0}^{N} |||\Phi_{\Omega}^{n}(f)|||^2 = \|f\|_2^2 - W_{N+1}(f) \tag{4.168}$$

$$\geqslant \int_{\mathbb{R}} |\widehat{f}(\omega)|^2 \Big|\widehat{r_l}\Big(\frac{\omega}{a^N\delta}\Big)\Big|^2 \mathrm{d}\omega \tag{4.169}$$

$$= \int_{B_L(0)} |\widehat{f}(\omega)|^2 \Big|\widehat{r_l}\Big(\frac{\omega}{a^N\delta}\Big)\Big|^2 \mathrm{d}\omega, \tag{4.170}$$

where (4.168) is by the lower bound in (4.98), (4.169) is thanks to Parseval's formula and (4.27) as well as (4.30), and (4.170) follows from f being L-band-limited. Next, thanks to $\widehat{r_l}$ monotonically decreasing in $|\omega|$, we get

$$\Big|\widehat{r_l}\Big(\frac{\omega}{a^N\delta}\Big)\Big|^2 \geqslant \Big|\widehat{r_l}\Big(\frac{L}{a^N\delta}\Big)\Big|^2, \qquad \forall\, \omega \in B_L(0). \tag{4.171}$$

Employing (4.171) in (4.170) yields

$$\|f\|_2^2 \geqslant \Big|\widehat{r_l}\Big(\frac{L}{a^N\delta}\Big)\Big|^2 \|f\|_2^2 = \Big(1 - \frac{L}{a^N\delta}\Big)_+^{2l} \|f\|_2^2 \tag{4.172}$$

$$= \Big(1 - \frac{L}{a^N\delta}\Big)^{2l} \|f\|_2^2 \geqslant (1-\varepsilon)\|f\|_2^2, \tag{4.173}$$

where in (4.172) we used Parseval's formula, the equality in (4.173) is by $L \leqslant a^N\delta$, which, in turn, is by (4.34), and the inequality in (4.173) is also due to (4.34) (upon rearranging terms). This establishes (4.32) and thereby completes the proof of ii).

4.7.8. Proof of Corollary 3

The proof is very similar to that of Corollary 2 in Section 4.7.7. We start with statement i). Let $f \in H^s(\mathbb{R}^d)\backslash\{0\}$ and $\varepsilon \in (0,1)$ and note that, by (4.87) and (4.88) together with $B_{\Omega}^N = 1$, $N \in \mathbb{N}$, which is by assumption, we have

$$W_N(f) \leqslant \frac{(2l)^{\gamma}\|f\|_{H^s}^2}{\delta^{\gamma} N^{\gamma\alpha}}, \quad \forall s > 0, \tag{4.174}$$

where $\gamma = \min\{1, 2s\}$. By Proposition 8 in Section 4.7.4 with $A_{\Omega}^N = B_{\Omega}^N = 1$, $N \in \mathbb{N}$, and $\lim_{N\to\infty} W_N(f) = 0$, $f \in L^2(\mathbb{R}^d)$, which follows from Proposition 9 in Section 4.7.5, we have

$$\begin{aligned}\|f\|_2^2 = |||\Phi_\Omega(f)|||^2 &= \sum_{n=0}^{\infty} |||\Phi_\Omega^n(f)|||^2 \\ &\geqslant \sum_{n=0}^{N} |||\Phi_\Omega^n(f)|||^2 = \|f\|_2^2 - W_{N+1}(f) && (4.175)\\ &\geqslant \|f\|_2^2 - \frac{(2l)^\gamma \|f\|_{H^s}^2}{\delta^\gamma (N+1)^{\gamma\alpha}} && (4.176)\\ &\geqslant \|f\|_2^2 - \varepsilon\|f\|_2^2 = (1-\varepsilon)\|f\|_2^2, && (4.177)\end{aligned}$$

where (4.175) is by the lower bound in (4.98), (4.176) is thanks to (4.174), and (4.177) follows from (4.35). This establishes (4.32) and thereby completes the proof of i).

The proof of statement ii) is very similar to that of statement i). Let $f \in H^s(\mathbb{R})\backslash\{0\}$ and $\varepsilon \in (0,1)$ and note that, by (4.129), (4.130), (4.153), and (4.154), we have

$$W_N(f) \leqslant \frac{(2l)^\gamma \|f\|_{H^s}^2}{\delta^\gamma a^{\gamma(N-1)}}, \quad \forall s > 0, \tag{4.178}$$

where $\gamma = \min\{1, 2s\}$. By Proposition 8 in Section 4.7.4 with $A_\Omega^N = B_\Omega^N = 1$, $N \in \mathbb{N}$, and $\lim_{N\to\infty} W_N(f) = 0$, $f \in L^2(\mathbb{R})$, which follows from Proposition 9 in Section 4.7.5, we have

$$\begin{aligned}\|f\|_2^2 = |||\Phi_\Omega(f)|||^2 &= \sum_{n=0}^{\infty} |||\Phi_\Omega^n(f)|||^2 \\ &\geqslant \sum_{n=0}^{N} |||\Phi_\Omega^n(f)|||^2 = \|f\|_2^2 - W_{N+1}(f) && (4.179)\\ &\geqslant \|f\|_2^2 - \frac{(2l)^\gamma \|f\|_{H^s}^2}{\delta^\gamma a^{\gamma N}} && (4.180)\\ &\geqslant \|f\|_2^2 - \varepsilon\|f\|_2^2 = (1-\varepsilon)\|f\|_2^2, && (4.181)\end{aligned}$$

where (4.179) is by the lower bound in (4.98), (4.180) is thanks to (4.178), and (4.181) follows from (4.36). This establishes (4.32) and thereby completes the proof of ii).

4.7.9. Proof of Corollary 4

Let a be the decay factor in (4.26) or (4.29). Then, it follows from (4.36) that

$$a \geqslant \left(\frac{2l \, \|f\|_{H^s}^{2/\gamma}}{\varepsilon^{1/\gamma} \delta \|f\|_2^{2/\gamma}} \right)^{1/N} = \kappa \tag{4.182}$$

is sufficient for (4.37) to hold. In the wavelet case, we have $a = \frac{r^2+1}{r^2-1}$, $r > 1$, which, when combined with (4.182), yields

$$\frac{r^2+1}{r^2-1} \geqslant \kappa. \tag{4.183}$$

Rearranging terms in (4.183) establishes (4.38). Next, in the Weyl-Heisenberg case, we have $a = \frac{1}{2} + \frac{\delta}{R}$, $\delta \geqslant \frac{R}{2}$, which, when combined with (4.182), leads to

$$\frac{1}{2} + \frac{\delta}{R} \geqslant \kappa. \tag{4.184}$$

Rearranging terms in (4.184) establishes (4.39) and thereby completes the proof.

CHAPTER 5

From theory to practice: Discrete-time deep convolutional neural networks

The first four chapters of this thesis focused on a mathematical theory of DCNNs for feature extraction in continuous time. This chapter considers the practically relevant discrete-time case, introduces new convolutional neural network architectures, and proposes a mathematical framework for their analysis. Specifically, we establish deformation and translation sensitivity results of local and global nature, and we investigate how certain structural properties of the input signal are reflected in the corresponding feature vectors. Our theory applies to general filters and general Lipschitz-continuous nonlinearities and pooling operators. For simplicity of exposition, we focus on the 1-D case throughout this chapter, noting that the extension to the higher-dimensional case does not pose any significant difficulties. Experiments on handwritten digit classification and facial landmark detection—including a feature importance evaluation—complement the theoretical findings.

Outline

The remainder of this chapter is organized as follows. Section 5.1 presents the notation and preparatory material of interest in the context of this chapter. In Section 5.2, we introduce the basic building blocks of the discrete-time DCNNs analyzed in this chapter, and in Section 5.3, we present the network topology. In Section 5.4, we define sampled cartoon functions which allow us to understand how certain structural properties of the input signal, such as the presence of sharp edges, are reflected in the feature vector. Section 5.5 contains our main results of this chapter, Theorems 5 and 6, which provide global and local feature vector properties, respectively. Finally, experiments on handwritten digit classification and facial landmark detection are presented in Section 5.6.

5.1. NOTATION AND PREPARATORY MATERIAL

We let $H_N := \{f : \mathbb{Z} \to \mathbb{C} \mid f[n] = f[n+N],\ \forall\, n \in \mathbb{Z}\}$ be the set of N-periodic discrete-time signals[1], and set $I_N := \{0, 1, \ldots, N-1\}$. The delta function $\delta \in H_N$ is $\delta[n] := 1$, for $n = kN$, $k \in \mathbb{Z}$, and $\delta[n] := 0$, else. For $f, g \in H_N$, we set $\langle f, g\rangle := \sum_{k\in I_N} f[k]\overline{g[k]}$, $\|f\|_1 := \sum_{n\in I_N} |f[n]|$, $\|f\|_2 := (\sum_{n\in I_N} |f[n]|^2)^{1/2}$, and $\|f\|_\infty := \sup_{n\in I_N} |f[n]|$. We denote the discrete Fourier transform (DFT) of $f \in H_N$ by $\hat{f}[k] := \sum_{n\in I_N} f[n]e^{-2\pi ikn/N}$. The circular convolution of $f \in H_N$ and $g \in H_N$ is $(f * g)[n] := \sum_{k\in I_N} f[k]g[n-k]$. We write $(T_m f)[n] := f[n-m]$, $m \in \mathbb{Z}$, for the cyclic translation operator. The supremum norm of a continuous-time function $c : \mathbb{R} \to \mathbb{C}$ is $\|c\|_\infty := \sup_{x\in\mathbb{R}} |c(x)|$.

[1]We note that H_N is isometrically isomorphic to $\mathbb{C}^N$, but we prefer to work with H_N for the sake of expositional simplicity.

5.2. THE BASIC BUILDING BLOCK

The basic building block of the discrete-time DCNNs we analyze in this chapter consists of a convolutional transform followed by a non-linearity and a pooling operator.

5.2.1. Convolutional transform

A convolutional transform is made up of a set of filters $\Psi_\Lambda = \{g_\lambda\}_{\lambda\in\Lambda}$. The finite index set Λ can be thought of as labeling a collection of scales, directions, or frequency-shifts. The filters g_λ—referred to as atoms—may be learned (in a supervised or unsupervised fashion), pre-specified and unstructured such as random filters, or pre-specified and structured such as wavelets, curvelets, shearlets, or Weyl-Heisenberg functions.

Definition 8. *Let Λ be a finite index set. The collection $\Psi_\Lambda = \{g_\lambda\}_{\lambda\in\Lambda} \subseteq H_N$ is called a convolutional set with Bessel bound $B \geqslant 0$ if*

$$\sum_{\lambda\in\Lambda} \|f * g_\lambda\|_2^2 \leqslant B\|f\|_2^2, \qquad \forall f \in H_N. \tag{5.1}$$

Condition (5.1) is equivalent to

$$\sum_{\lambda\in\Lambda} |\widehat{g_\lambda}[k]|^2 \leqslant B, \qquad \forall k \in I_N, \tag{5.2}$$

and hence, *every* finite set $\{g_\lambda\}_{\lambda\in\Lambda}$ is a convolutional set with Bessel bound $B^* := \max_{k\in I_N} \sum_{\lambda\in\Lambda} |\widehat{g_\lambda}[k]|^2$. As $(f * g_\lambda)[n] = \langle f, \overline{g_\lambda[n-\cdot]}\rangle$, $n \in I_N$, $\lambda \in \Lambda$, the outputs of the filters g_λ may be interpreted as inner products of the input signal f with translates of the atoms g_λ. Frame theory (Daubechies, 1992) therefore tells us that the existence of a lower bound $A > 0$ in (5.2) according to

$$A \leqslant \sum_{\lambda\in\Lambda} |\widehat{g_\lambda}[k]|^2 \leqslant B, \qquad \forall k \in I_N, \tag{5.3}$$

implies that every element in H_N can be written as a linear combination of elements in the set $\left\{\overline{g_\lambda[n-\cdot]}\right\}_{n\in I_N,\lambda\in\Lambda}$ (or in more technical

parlance, the set $\{\overline{g_\lambda[n-\cdot]}\}_{n\in I_N,\lambda\in\Lambda}$ is complete for H_N). The absence of a lower bound $A > 0$ may therefore result in Ψ_Λ failing to extract essential features of the signal f. We note, however, that even learned filters are likely to satisfy (5.3) as all that is needed is, for each $k \in I_N$, to have $\widehat{g_\lambda}[k] \neq 0$ for at least one $\lambda \in \Lambda$. As we shall see below, the existence of a lower bound $A > 0$ in (5.3) is, however, not needed for our theory to apply.

Examples of structured convolutional sets with $A = B = 1$ include, in the 1-D case, wavelets (Daubechies, 1992) and Weyl-Heisenberg functions (Bölcskei and Hlawatsch, 1997), and in the 2-D case, tensorized wavelets (Mallat, 2009), curvelets (Candès et al., 2006), and shearlets (Kutyniok and Labate, 2012b).

5.2.2. Non-linearities

The non-linearities $\rho : \mathbb{C} \to \mathbb{C}$ we consider are all point-wise and satisfy the Lipschitz property $|\rho(x) - \rho(y)| \leqslant L|x - y|$, $\forall x, y \in \mathbb{C}$, for some $L > 0$.

Example non-linearities

i) The *hyperbolic tangent* non-linearity, defined as

$$\rho(x) = \tanh(\mathrm{Re}(x)) + i \tanh(\mathrm{Im}(x)),$$

where $\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$, has Lipschitz constant $L = 2$.

ii) The *rectified linear unit* non-linearity is given by

$$\rho(x) = \max\{0, \mathrm{Re}(x)\} + i \max\{0, \mathrm{Im}(x)\},$$

and has Lipschitz constant $L = 2$.

iii) The *modulus* non-linearity is $\rho(x) = |x|$, and has Lipschitz constant $L = 1$.

iv) The logistic sigmoid non-linearity is defined as

$$\rho(x) = \mathrm{sig}(\mathrm{Re}(x)) + i \,\mathrm{sig}(\mathrm{Im}(x)),$$

where $\mathrm{sig}(x) = \frac{1}{1+e^{-x}}$, and has Lipschitz constant $L = 1/2$.

We refer the reader to Section 2.3 for proofs of the Lipschitz properties of these example non-linearities.

5.2.3. Pooling operators

The essence of pooling is to reduce signal dimensionality in the individual network layers and to ensure robustness of the feature vector w.r.t. deformations and translations.

The theory developed in this chapter applies to general pooling operators $P : H_N \to H_{N/S}$, where $N, S \in \mathbb{N}$ with $N/S \in \mathbb{N}$, that satisfy the Lipschitz property $\|Pf - Pg\|_2 \leqslant R\|f - g\|$, $\forall f, g \in H_N$, for some $R > 0$. The integer S will be referred to as pooling factor, and determines the "size" of the neighborhood values are combined in, see Fig. 5.1 for an illustrative example.

Example pooling operators

i) *Sub-sampling*, defined as $P : H_N \to H_{N/S}$,

$$(Pf)[n] = f[Sn], \quad n \in I_{N/S},$$

has Lipschitz constant $R = 1$. For $S = 1$, P is the identity operator which amounts to "no pooling".

ii) *Averaging*, defined as $P : H_N \to H_{N/S}$,

$$(Pf)[n] = \sum_{k=Sn}^{Sn+S-1} \alpha_{k-Sn} f[k], \quad n \in I_{N/S},$$

has Lipschitz constant $R = S^{1/2} \max_{k\in\{0,\ldots,S-1\}} |\alpha_k|$. The weights $\{\alpha_k\}_{k=0}^{S-1}$ can be learned (LeCun et al., 1998) or pre-specified (Pinto et al., 2008) (e.g., uniform pooling corresponds to $\alpha_k = \frac{1}{S}$, for $k \in \{0, \ldots, S-1\}$).

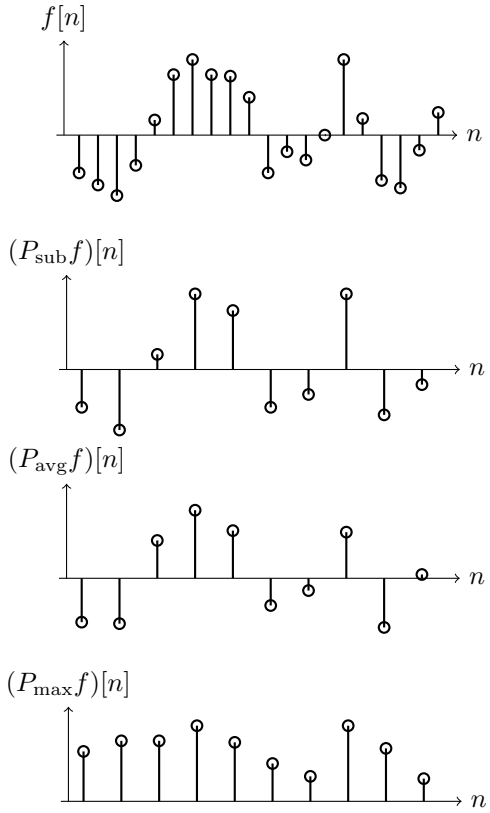

Fig. 5.1: Impact of pooling operators (with $S = 2$) on the signal $f \in H_{20}$ (top row). Pooling by sub-sampling amounts to retaining every second sample. Pooling by averaging amounts to computing local averages of two consecutive samples. Pooling by maximization amounts to picking the maximal value of two consecutive samples. Here, we used the notation sub.: sub-sampling, avg.: average-pooling, and max.: max-pooling.

iii) *Maximization*, defined as $P : H_N \to H_{N/S}$,

$$(Pf)[n] = \max_{k \in \{Sn,\ldots,Sn+S-1\}} |f[k]|, \quad n \in I_{N/S},$$

has Lipschitz constant $R = 1$.

We refer to Section 5.7.1 for proofs of the Lipschitz property of these three example pooling operators along with the derivations of the corresponding Lipschitz constants.

5.3. THE NETWORK ARCHITECTURE

The architecture we consider is flexible in the following sense. In each layer, we can feed into the feature vector either the signals propagated down to that layer (i.e., the feature maps), filtered versions thereof, or we can decide not to have that layer contribute to the feature vector.

The basic building blocks of our network are the triplets (Ψ_d, ρ_d, P_d) of filters, non-linearities, and pooling operators associated with the d-th network layer and referred to as *modules*. We emphasize that these triplets are allowed to be different across layers.

Definition 9. *For network layers d, $1 \leqslant d \leqslant D$, let $\Psi_d = \{g_{\lambda_d}\}_{\lambda_d \in \Lambda_d} \subseteq H_{N_d}$ be a convolutional set, $\rho_d : \mathbb{C} \to \mathbb{C}$ a point-wise Lipschitz-continuous non-linearity, and $P_d : H_{N_d} \to H_{N_{d+1}}$ a Lipschitz-continuous pooling operator with $N_{d+1} = \frac{N_d}{S_d}$, where $S_d \in \mathbb{N}$ denotes the pooling factor in the d-th layer. Then, the sequence of triplets*

$$\Omega := \Big(\big(\Psi_d, \rho_d, P_d \big) \Big)_{1 \leqslant d \leqslant D}$$

is called a module-sequence.

Note that the dimensions of the spaces H_{N_d} satisfy $N_1 \geqslant N_2 \geqslant \ldots \geqslant N_D$. Associated with the module (Ψ_d, ρ_d, P_d), we define the operator

$$(U_d[\lambda_d] f) := P_d(\rho_d(f * g_{\lambda_d})) \tag{5.4}$$

and extend it to paths on index sets

$$q = (\lambda_1, \lambda_2, \ldots, \lambda_d) \in \Lambda_1 \times \Lambda_2 \times \cdots \times \Lambda_d := \Lambda^d,$$

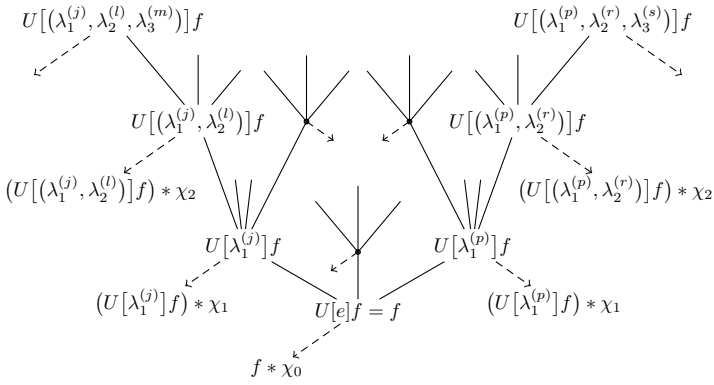

Fig. 5.2: Network architecture underlying the feature extractor (5.6). The index $\lambda_d^{(k)}$ corresponds to the k-th atom $g_{\lambda_d^{(k)}}$ of the convolutional set Ψ_d associated with the d-th network layer. The function χ_d is the output-generating atom of the d-th layer. The root of the network corresponds to $d = 0$.

for $1 \leqslant d \leqslant D$, according to

$$U[q]f = U[(\lambda_1, \lambda_2, \ldots, \lambda_d)]f := U_d[\lambda_d] \cdots U_2[\lambda_2] U_1[\lambda_1] f. \qquad (5.5)$$

For the empty path $e := \emptyset$ we set $\Lambda^0 := \{e\}$ and let $U[e]f := f$, for all $f \in H_{N_1}$.

The network output in the d-th layer is given by $(U[q]f) * \chi_d$, $q \in \Lambda^d$, where $\chi_d \in H_{N_{d+1}}$ is referred to as output-generating atom. Specifically, we let χ_d be (i) the delta function $\delta[n]$, $n \in I_{N_{d+1}}$, if we want the output to equal the unfiltered features $U[q]f$, $q \in \Lambda^d$, propagated down to layer d, or (ii) any other signal of length N_{d+1}, or (iii) $\chi_d = 0$ if we do not want layer d to contribute to the feature vector. From now on we formally add χ_d to the set $\Psi_{d+1} = \{g_{\lambda_{d+1}}\}_{\lambda_{d+1} \in \Lambda_{d+1}}$, noting that $\{g_{\lambda_{d+1}}\}_{\lambda_{d+1} \in \Lambda_{d+1}} \cup \{\chi_d\}$ forms a convolutional set Ψ'_{d+1} with Bessel bound $B'_{d+1} \leqslant B_{d+1} + \max_{k \in I_{N_{d+1}}} |\widehat{\chi}_d[k]|^2$. We emphasize that the atoms of the augmented set $\{g_{\lambda_{d+1}}\}_{\lambda_{d+1} \in \Lambda_{d+1}} \cup \{\chi_d\}$ are employed across two consecutive layers in the sense of χ_d generating

the output in the d-th layer according to $(U[q]f) * \chi_d$, $q \in \Lambda^d$, and the remaining atoms $\{g_{\lambda_{d+1}}\}_{\lambda_{d+1} \in \Lambda_{d+1}}$ propagating the signals $U[q]f$, $q \in \Lambda^d$, from the d-th layer down to the $(d+1)$-st layer according to (5.4), see Fig. 5.2. With slight abuse of notation, we shall henceforth write Ψ_d for Ψ'_d and B_d for B'_d as well.

We are now ready to define the feature extractor Φ_Ω based on the module-sequence Ω.

Definition 10. *Let $\Omega = \big((\Psi_d, \rho_d, P_d)\big)_{1 \leqslant d \leqslant D}$ be a module-sequence. The feature extractor Φ_Ω based on Ω maps $f \in H_{N_1}$ to its features*

$$\Phi_\Omega(f) := \bigcup_{d=0}^{D-1} \Phi_\Omega^d(f), \tag{5.6}$$

where $\Phi_\Omega^d(f) := \{(U[q]f) * \chi_d\}_{q \in \Lambda^d}$ is the collection of features generated in the d-th network layer (see Fig. 5.2).

The dimension of the feature vector $\Phi_\Omega(f)$ is given by

$$\varepsilon_0 N_1 + \sum_{d=1}^{D-1} \varepsilon_d N_{d+1} \big(\prod_{k=1}^{d} \mathrm{card}(\Lambda_k) \big),$$

where $\varepsilon_d = 1$, if an output is generated (either filtered or unfiltered) in the d-th network layer, and $\varepsilon_d = 0$, else. As $N_{d+1} = \frac{N_d}{S_d} = \cdots = \frac{N_1}{S_1 \cdots S_d}$, for $d \geqslant 1$, the dimension of the overall feature vector is determined by the pooling factors S_k and, of course, the layers that contribute to the feature vector.

Remark 8. *It was argued in (Bruna and Mallat, 2013; Andén and Mallat, 2014; Oyallon and Mallat, 2015) that the features $\Phi_\Omega^1(f)$ when generated by wavelet filters, modulus non-linearities, without intra-layer pooling, and by employing output-generating atoms with low-pass characteristics, describe mel frequency cepstral coefficients (Davis and Mermelstein, 1980) in 1-D, and SIFT-descriptors (Lowe, 2004; Tola et al., 2010) in 2-D.*

5.4. SAMPLED CARTOON FUNCTIONS

While our main results hold for general signals f, we can provide a refined analysis for the class of sampled cartoon functions. This allows to understand how certain structural properties of the input signal, such as the presence of sharp edges, are reflected in the feature vector. As already mentioned in Section 3.4.3, cartoon functions—as introduced in continuous time in (Donoho, 2001)—are piecewise "smooth" apart from curved discontinuities along C^2-hypersurfaces. They hence provide a good model for natural images (see Fig. 3.7, left) such as those in the Caltech-256 (Griffin et al., 2007) and the CIFAR-100 (Krizhevsky, 2009) data sets, for images of handwritten digits (LeCun and Cortes, 1998) (see Fig. 3.7, right), and for images of geometric objects of different shapes, sizes, and colors as in the Baby AI School data set[2].

We refer the reader to Section 3.4.3 for bounds on deformation sensitivity for cartoon functions in continuous time DCNNs. Here, we analyze deformation sensitivity for sampled cartoon functions passed through discrete-time DCNNs.

Definition 11. *The function $c : \mathbb{R} \to \mathbb{C}$ is referred to as a cartoon function if it can be written as $c = c_1 + \mathbb{1}_{[a,b]} c_2$, where $[a, b] \subseteq [0, 1]$ is a closed interval, and $c_i : \mathbb{R} \to \mathbb{C}$, $i = 1, 2$, satisfies the Lipschitz property*[3]

$$|c_i(x) - c_i(y)| \leqslant C|x - y|, \quad \forall x, y \in \mathbb{R}, \tag{5.8}$$

[2] `http://www.iro.umontreal.ca/%7Elisa/twiki/bin/view.cgi/Public/BabyAISchool`

[3] We note that it is actually the condition

$$|\nabla c_i(x)| \leq C\langle x\rangle^{-1}, \quad i = 1, 2, \tag{5.7}$$

for some $C > 0$, rather than (5.8) that was introduced in Definition 7 in Section 3.4.3. In this chapter, however, we prefer to work with condition (5.8), which is less restrictive than (5.7) due to the fact that every continuously differentiable function with bounded derivative is Lipschitz-continuous, see, e.g., (Searcóid, 2007, Theorem 9.5.1).

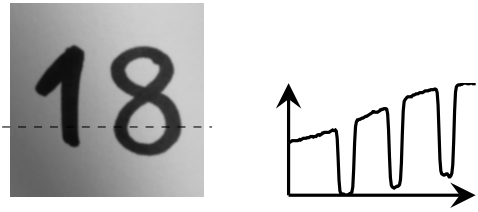

Fig. 5.3: Left: Image of a handwritten digit. Right: Pixel values corresponding to the dashed row in the left image.

for some $C > 0$. Furthermore, we denote by

$$\mathcal{C}^K_{\mathrm{CART}} := \{c_1 + \mathbb{1}_{[a,b]}c_2 \mid |c_i(x) - c_i(y)| \leqslant K|x-y|, \\ \forall\, x, y \in \mathbb{R},\ i = 1,2,\ \|c_2\|_\infty \leqslant K\}$$

the class of cartoon functions of variation $K > 0$, and by

$$\mathcal{C}^{N,K}_{\mathrm{CART}} := \Big\{ f[n] = c(n/N),\ n \in \{0,1,\ldots,N-1\} \ \Big| \\ c = (c_1 + \mathbb{1}_{[a,b]}c_2) \in \mathcal{C}^K_{\mathrm{CART}} \text{ with } a, b \notin \Big\{0, \frac{1}{N}, \ldots, \frac{N-1}{N}\Big\}\Big\}$$

the class of sampled cartoon functions of length N and variation $K > 0$.

We note that excluding the boundary points a, b of the interval $[a, b]$ from being sampling points n/N in the definition of $\mathcal{C}^{N,K}_{\mathrm{CART}}$ is of conceptual importance (see Remark 11 in the Section 5.7.3). Moreover, our results can easily be generalized to classes $\mathcal{C}^{N,K}_{\mathrm{CART}}$ consisting of functions $f[n] = c(n/N)$ with c containing multiple "1-D edges" (i.e., multiple discontinuity points) according to $c = c_1 + \sum_{l=1}^{L} \mathbb{1}_{[a_l,b_l]}c_2$ with $\cap_{l=1}^{L}[a_l, b_l] = \emptyset$. We also note that $\mathcal{C}^{N,K}_{\mathrm{CART}}$ reduces to the class of sampled Lipschitz-continuous functions upon setting $c_2 = 0$.

A sampled cartoon function in 2-D models, e.g., an image acquired by a digital camera (see Fig. 5.3, left); in 1-D, $f \in \mathcal{C}^{N,K}_{\mathrm{CART}}$ can be thought of as the pixels in a row or column of this image (see Fig. 5.3 right, which shows a cartoon function with 6 discontinuity points).

5.5. ANALYTICAL RESULTS

We analyze global and local feature vector properties with globality pertaining to characteristics brought out by the union of features across all network layers, and locality identifying attributes made explicit in individual layers.

5.5.1. Global properties

Theorem 5. *Let* $\Omega = \big((\Psi_d, \rho_d, P_d)\big)_{1 \leqslant d \leqslant D}$ *be a module-sequence. Assume that the Bessel bounds* $B_d > 0$*, the Lipschitz constants* $L_d > 0$ *of the non-linearities* ρ_d*, and the Lipschitz constants* $R_d > 0$ *of the pooling operators* P_d *satisfy*

$$\max_{1 \leqslant d \leqslant D} \max\{B_d, B_d R_d^2 L_d^2\} \leqslant 1. \tag{5.9}$$

i) The feature extractor Φ_Ω *is Lipschitz-continuous with Lipschitz constant* $L_\Omega = 1$*, i.e.,*

$$|||\Phi_\Omega(f) - \Phi_\Omega(h)||| \leqslant \|f - h\|_2, \tag{5.10}$$

for all $f, h \in H_{N_1}$*, where the feature space norm is defined as*

$$|||\Phi_\Omega(f)|||^2 := \sum_{d=0}^{D-1} \sum_{q \in \Lambda^d} ||(U[q]f) * \chi_d||_2^2. \tag{5.11}$$

ii) If, in addition to (5.9)*, for all* $d \in \{1, \ldots, D-1\}$ *the non-linearities* ρ_d *and the pooling operators* P_d *satisfy* $\rho_d(0) = 0$ *and* $P_d(0) = 0$ *(as all non-linearities and pooling operators in the Sections 5.2.2 and 5.2.3, apart from the logistic sigmoid non-linearity, do), then*

$$|||\Phi_\Omega(f)||| \leqslant \|f\|_2, \quad \forall f \in H_{N_1}. \tag{5.12}$$

iii) For every variation $K > 0$ *and deformation* F_τ *of the form*

$$(F_\tau f)[n] := c(n/N_1 - \tau(n/N_1)), \quad n \in I_{N_1}, \tag{5.13}$$

where $\tau : \mathbb{R} \to [-1, 1]$*, the deformation sensitivity is bounded according to*

$$|||\Phi_\Omega(F_\tau f) - \Phi_\Omega(f)||| \leqslant 4KN_1^{1/2}\|\tau\|_\infty^{1/2}, \qquad (5.14)$$

for all $f \in \mathcal{C}_{\mathrm{CART}}^{N_1,K}$.

Proof. The proof is given in Section 5.7.2. □

The Lipschitz continuity (5.10) guarantees that pairwise distances of input signals do not increase through feature extraction. As an immediate implication of the Lipschitz continuity we get robustness of the feature extractor w.r.t. additive bounded noise $\eta \in H_{N_1}$ in the sense of

$$|||\Phi_\Omega(f + \eta) - \Phi_\Omega(f)||| \leqslant \|\eta\|_2,$$

for all $f \in H_{N_1}$.

Remark 9. *As detailed in the proof of Theorem 5, the Lipschitz continuity* (5.10) *combined with the deformation sensitivity bound (see Proposition 10 in the Section 5.7.3) for the signal class under consideration, namely sampled cartoon functions, establishes the deformation sensitivity bound* (5.14) *for the feature extractor. This insight has important practical ramifications as it shows that whenever we have deformation sensitivity bounds for a signal class, we automatically get deformation sensitivity guarantees for the corresponding feature extractor.*

From (5.14) we can deduce a statement on the sensitivity of Φ_Ω w.r.t. translations on $\mathbb{R}$. To this end, we first note that setting $\tau_t(x) = t$, $x \in \mathbb{R}$, for $t \in [-1, 1]$, (5.13) becomes

$$(F_{\tau_t} f)[n] = c(n/N_1 - t), \quad n \in I_{N_1}.$$

Particularizing (5.14) accordingly, we obtain

$$|||\Phi_\Omega(F_{\tau_t} f) - \Phi_\Omega(f)||| \leqslant 4KN_1^{1/2}|t|^{1/2}, \qquad (5.15)$$

which shows that small translations $|t|$ of the underlying analog signal $c(x)$, $x \in \mathbb{R}$, lead to small changes in the feature vector obtained by

passing the resulting sampled signal through a discrete-time DCNN. We shall say that (5.15) is a translation sensitivity bound. Analyzing the impact of deformations and translations over $\mathbb{R}$ on the discrete feature vector generated by the sampled analog signal closely models real-world phenomena (e.g., the jittered acquisition of an analog signal with a digital camera, where different values of N_1 in (5.13) correspond to different camera resolutions).

We note that, while iii) in Theorem 5 is specific to cartoon functions, i) and ii) apply to all signals in H_{N_1}.

The strength of the results in Theorem 5 derives itself from the fact that condition (5.9) on the underlying module-sequence Ω is easily met in practice. To see this, we first note that B_d is determined by the convolutional set Ψ_d, L_d by the non-linearity ρ_d, and R_d by the pooling operator P_d. Condition (5.9) is met if

$$B_d \leqslant \min\{1, R_d^{-2} L_d^{-2}\}, \quad \forall\, d \in \{1, 2, \ldots, D\}, \tag{5.16}$$

which, if not satisfied by default, can be enforced simply by normalizing the elements in Ψ_d. Specifically, for $C_d := \max\{B_d, R_d^2 L_d^2\}$ the set $\widetilde{\Psi}_d := \{C_d^{-1/2} g_{\lambda_d}\}_{\lambda_d \in \Lambda_d}$ has Bessel bound $\widetilde{B_d} = \frac{B_d}{C_d}$ and hence satisfies (5.16). While this normalization does not have an impact on the results in Theorem 5, there exists, however, a tradeoff between energy preservation and deformation (respectively translation) sensitivity in Φ_Ω^d as detailed in Section 5.5.2.

5.5.2. Local properties

Theorem 6. *Let $\Omega = \big((\Psi_d, \rho_d, P_d)\big)_{1 \leqslant d \leqslant D}$ be a module-sequence with corresponding Bessel bounds $B_d > 0$, Lipschitz constants $L_d > 0$ of the non-linearities ρ_d, Lipschitz constants $R_d > 0$ of the pooling operators P_d, and output-generating atoms χ_d. Let further $L_\Omega^0 := \|\chi_0\|_1$ and*[4]

$$L_\Omega^d := \|\chi_d\|_1 \Big(\prod_{k=1}^{d} B_k L_k^2 R_k^2 \Big)^{1/2}, \quad d \geqslant 1. \tag{5.17}$$

[4]We note that $\|\chi_d\|_1$ in (5.17) can be upper-bounded (and hence substituted) by $\sqrt{B_{d+1}}$, see Remark 13 in Section 5.7.4.

i) The features generated in the d-th network layer are Lipschitz-continuous with Lipschitz constant L_Ω^d, i.e.,

$$|||\Phi_\Omega^d(f) - \Phi_\Omega^d(h)||| \leqslant L_\Omega^d \|f - h\|_2, \tag{5.18}$$

for all $f, h \in H_{N_1}$, where $|||\Phi_\Omega^d(f)|||^2 := \sum_{q\in\Lambda^d} ||(U[q]f) * \chi_d||_2^2$.

ii) If the non-linearities ρ_k and the pooling operators P_k satisfy $\rho_k(0) = 0$ and $P_k(0) = 0$, respectively, for all $k \in \{1, \dots, d\}$, then

$$|||\Phi_\Omega^d(f)||| \leqslant L_\Omega^d \|f\|_2, \quad \forall f \in H_{N_1}. \tag{5.19}$$

iii) For all $K > 0$ and all $\tau : \mathbb{R} \to [-1, 1]$, the features generated in the d-th network layer satisfy

$$|||\Phi_\Omega^d(F_\tau f) - \Phi_\Omega^d(f)||| \leqslant 4L_\Omega^d K N^{1/2} \|\tau\|_\infty^{1/2}, \tag{5.20}$$

for all $f \in \mathcal{C}_{\mathrm{CART}}^{N_1,K}$, where $F_\tau f$ is defined in (5.13).

iv) If the module-sequence employs sub-sampling, average pooling, or max-pooling with corresponding pooling factors $S_d \in \mathbb{N}$, then

$$\Phi_\Omega^d(T_m f) = T_{\frac{m}{S_1 \cdots S_d}} \Phi_\Omega^d(f), \tag{5.21}$$

for all $f \in H_{N_1}$ and all $m \in \mathbb{Z}$ with $\frac{m}{S_1 \cdots S_d} \in \mathbb{Z}$. Here, $T_m \Phi_\Omega^d(f)$ refers to element-wise application of T_m, i.e.,

$$T_m \Phi_\Omega^d(f) := \{T_m h \mid h \in \Phi_\Omega^d(f)\}.$$

Proof. The proof is given in Section 5.7.4. □

One may be tempted to infer the global results (5.10), (5.12), and (5.14) in Theorem 5 in Section 5.5.1 from the corresponding local results in Theorem 6, e.g., the energy bound in (5.12) from (5.19) according to

$$|||\Phi_\Omega(f)||| = \Big(\sum_{d=0}^{D-1} |||\Phi_\Omega^d(f)|||^2 \Big)^{1/2} \leqslant \sqrt{D} \|f\|_2,$$

where we employed $L_\Omega^d \leqslant 1$ owing to (5.9). This would, however, lead to the "global" Lipschitz constant $L_\Omega = 1$ in (5.10), (5.12), and (5.14) to be replaced by $L_\Omega = \sqrt{D}$ and thereby render the corresponding results much weaker.

Again, we emphasize that, while iii) in Theorem 6 is specific to cartoon functions, i), ii), and iv) apply to all signals in H_{N_1}.

For a fixed network layer d, the "local" Lipschitz constant L_Ω^d determines the noise sensitivity of the features $\Phi_\Omega^d(f)$ according to

$$|||\Phi_\Omega^d(f+\eta) - \Phi_\Omega^d(f)||| \leqslant L_\Omega^d \|\eta\|_2, \tag{5.22}$$

where (5.22) follows from (5.18). Moreover, L_Ω^d via (5.20) also quantifies the impact of deformations (or translations when $\tau_t(x) = t$, $x \in \mathbb{R}$, for $t \in [-1, 1]$) on the feature vector. In practice, it may be desirable to have the features Φ_Ω^d become more robust to additive noise and less deformation-sensitive (respectively, translation-sensitive) as we progress deeper into the network. Formally, this vertical sensitivity reduction can be induced by ensuring that $L_\Omega^{d+1} < L_\Omega^d$. Thanks to

$$L_\Omega^d = \frac{\|\chi_d\|_1 B_d^{1/2} L_d R_d}{\|\chi_{d-1}\|_1} L_\Omega^{d-1},$$

this can be accomplished by choosing the module-sequence such that $\|\chi_d\|_1 B_d^{1/2} L_d R_d < \|\chi_{d-1}\|_1$. Note, however, that owing to (5.19) this will also reduce the signal energy contained in the features $\Phi_\Omega^d(f)$. We therefore have a tradeoff between deformation (respectively translation) sensitivity and energy preservation. Having control over this tradeoff through the choice of the module-sequence Ω may come in handy in practice.

For average pooling with uniform weights $\alpha_k^d = \frac{1}{S_d}$, $k = 0, \ldots, S_d - 1$ (noting that the corresponding Lipschitz constant is $R_d = S_d^{-1/2}$, see Section 5.2.3), we get

$$L_\Omega^d = \|\chi_d\|_1 \Big(\prod_{k=1}^{d} \frac{B_k L_k^2}{S_k} \Big)^{1/2},$$

which illustrates that pooling can have an impact on the sensitivity and energy properties of Φ_Ω^d.

We finally turn to interpreting the translation covariance result (5.21). Owing to the condition $\frac{m}{S_1 \dots S_d} \in \mathbb{Z}$, we get translation covariance only on the rough grid induced by the product of the pooling factors. In the absence of pooling, i.e., $S_k = 1$, for $k \in \{1, \dots, d\}$, we obtain translation covariance w.r.t. the fine grid the input signal $f \in H_{N_1}$ lives on.

Remark 10. *We note that ScatNets (Bruna and Mallat, 2013) are translation-covariant on the rough grid induced by the factor 2^J corresponding to the coarsest wavelet scale. Our result in* (5.21) *is hence in the spirit of (Bruna and Mallat, 2013) with the difference that the grid in our case is induced by the pooling factors S_k.*

5.6. EXPERIMENTS

[5]We consider the problem of handwritten digit classification and evaluate the performance of the feature extractor Φ_Ω in combination with a SVM. The results we obtain are competitive with the state-of-the-art in the literature. The second line of experiments we perform assesses the importance of the features extracted by Φ_Ω in facial landmark detection and in handwritten digit classification, using random forests (RF) for regression and classification, respectively. Our results are based on a DCNN with different non-linearities and pooling operators, and with tensorized (i.e., separable) wavelets as filters, sensitive to 3 directions (horizontal, vertical, and diagonal). Furthermore, we generate outputs in all layers through low-pass filtering. Circular convolutions with the 1-D filters underlying the tensorized wavelets are efficiently implemented using the *algorithme à trous* (Holschneider et al., 1989).

To reduce the dimension of the feature vector, we compute features along frequency decreasing paths only (Bruna and Mallat, 2013), i.e.,

[5]Code available at `http://www.nari.ee.ethz.ch/commth/research/`

for every node $U[q]f$, $q \in \Lambda_1^{d-1}$, we retain only those child nodes $U_d[\lambda_d]U[q]f = P_d\big(\rho_d((U[q]f) * g_{\lambda_d})\big)$ that correspond to wavelets g_{λ_d} with scales larger than the maximum scale of the wavelets used to get $U[q]f$. We refer to (Bruna and Mallat, 2013; Waldspurger, 2017) for a detailed justification of this approach for scattering networks.

5.6.1. Handwritten digit classification

We use the MNIST data set of handwritten digits (LeCun and Cortes, 1998) which comprises 60,000 training and 10,000 test images of size 28×28. We set $D = 3$, and compare different network configurations, each defined by a single module (i.e., we use the same filters, non-linearity, and pooling operator in all layers). Specifically, we consider Haar wavelets and reverse biorthogonal 2.2 (RBIO2.2) wavelets (Mallat, 2009), both with $J = 3$ scales, the non-linearities described in Section 5.2.2, and the pooling operators described in Section 5.2.3 (with $S_1 = 1$ and $S_2 = 2$). We use a SVM with radial basis function (RBF) kernel for classification. To reduce the dimension of the feature vectors from 18,424 (or 50,176, for the configurations without pooling) down to 1000, we employ the supervised orthogonal least squares feature selection procedure described in (Oyallon and Mallat, 2015). The penalty parameter of the SVM and the localization parameter of the RBF kernel are selected via 10-fold cross-validation for each combination of wavelet filter, non-linearity, and pooling operator.

Table 5.1 shows the resulting classification errors on the test set. Configurations employing RBIO2.2 wavelets tend to yield a marginally lower classification error than those using Haar wavelets. For the tanh and LogSig non-linearities, max-pooling leads to a considerably lower classification error than other pooling operators. The configurations involving the modulus and ReLU non-linearities achieve classification accuracy competitive with the state-of-the-art (Bruna and Mallat, 2013) (class. err.: 0.43%), which is based on directional non-separable wavelets with 6 directions without intra-layer pooling. This is interesting as the separable wavelet filters employed here can be implemented more efficiently.

	Haar				RBIO2.2			
	abs	ReLU	tanh	LogSig	abs	ReLU	tanh	LogSig
n.p.	0.55	0.57	1.41	1.49	0.50	0.54	1.01	1.18
sub.	0.60	0.58	1.25	1.45	0.59	0.62	1.04	1.13
max.	0.61	0.60	0.68	0.76	0.55	0.56	0.71	0.75
avg.	0.57	0.58	1.26	1.44	0.51	0.60	1.04	1.18

Table 5.1: Classification error in percent for handwritten digit classification using different configurations of wavelet filters, non-linearities, and pooling operators (sub.: sub-sampling; max.: max-pooling; avg.: average-pooling; n.p.: no pooling).

5.6.2. Feature importance evaluation

In this experiment, we investigate the "importance" of the features generated by Φ_Ω corresponding to different layers, wavelet scales, and directions in two different learning tasks, namely, facial landmark detection and handwritten digit classification. The primary goal of this experiment is to illustrate the practical relevance of the notion of local properties of Φ_Ω as established in Section 5.5.2. For facial landmark detection we employ a RF regressor and for handwritten digit classification a RF classifier (Breiman, 2001). In both cases, we fix the number of trees to 30 and select the tree depth using out-of-bag error estimates (noting that increasing the number of trees does not significantly increase the accuracy). The impurity measure used for learning the node tests is the mean square error for facial landmark detection and the Gini impurity for handwritten digit classification. In both cases, feature importance is assessed using the Gini importance (Breiman et al., 1984), averaged over all trees. The Gini importance $I(\theta, T)$ of feature θ in the (trained) tree T is defined as

$$I(\theta, T) = \sum_{\ell \in T\colon \varphi(\ell)=\theta} \frac{n_\ell}{n_{\text{tot}}} \Big(\hat{\imath}_\ell - \frac{n_{\ell_L}}{n_\ell} \hat{\imath}_{\ell_L} - \frac{n_{\ell_R}}{n_\ell} \hat{\imath}_{\ell_R} \Big),$$

where $\varphi(\ell)$ denotes the feature determined in the training phase for the test at node ℓ, n_ℓ is the number of training samples passed through node ℓ, $n_{\text{tot}} = \sum_{\ell \in T} n_\ell$, $\hat{\imath}_\ell$ is the impurity at node ℓ, and ℓ_L and

Fig. 5.4: Images from the Caltech 10,000 Web Faces data base (Angelova et al., 2005) with corresponding annotations for eyes, nose, and mouth.

ℓ_R denote the left and right child node, respectively, of node ℓ. For the feature extractor Φ_Ω we set $D = 4$, employ Haar wavelets with $J = 3$ scales and the modulus non-linearity in every network layer, no pooling in the first layer and average pooling with uniform weights $1/S_d^2$, $S_d = 2$, in layers $d = 2, 3$.

Facial landmark detection

We use the Caltech 10,000 Web Faces data base (Angelova et al., 2005). Each of the 7092 images in the data base depicts one or more faces in different contexts (e.g., portrait images, groups of people), see Fig. 5.4. The data base contains annotations of the positions of eyes, nose, and mouth for at least one face per image. The learning task is to estimate the positions of these facial landmarks. The annotations serve as ground truth for training and testing. We preprocess the data set as follows. The patches containing the faces are extracted from the images using the Viola-Jones face detector (Viola and Jones, 2004). After discarding false positives, the patches are converted to grayscale and resampled to size 120×120 (using linear interpolation), before feeding them to the feature extractor Φ_Ω. This procedure yields a data set containing a total of 8776 face images. We select 80% of the images uniformly at random to form a training set and use the remaining images for testing. We train a separate RF for each facial landmark. Following (Dantone et al., 2012) we report the localization error, i.e., the ℓ_2-distance between the estimated and the ground

	left eye	right eye	nose	mouth	digits	disp. digits
Layer 0	0.020	0.023	0.016	0.014	0.046	0.004
Layer 1	0.629	0.646	0.576	0.490	0.426	0.094
Layer 2	0.261	0.236	0.298	0.388	0.337	0.280
Layer 3	0.090	0.095	0.110	0.108	0.192	0.622

Table 5.2: Cumulative feature importance per layer. Columns 1–4: facial landmark detection. Columns 5 and 6: handwritten digit classification.

truth landmark positions, on the test set as a fraction of the (true) inter-ocular distance. The errors obtained are: left eye: 0.062; right eye: 0.064; nose; 0.080, mouth: 0.095. As an aside, we note that these values are comparable with the ones reported in (Dantone et al., 2012) for a conditional RF using patch comparison features (evaluated on a different data set and a larger set of facial landmarks).

Handwritten digit classification

For this experiment, we again rely on the MNIST data set. The training set is obtained by sampling uniformly at random 1,000 images per digit from the MNIST training data set and we use the complete MNIST test set. We train two RFs, one based on unmodified images, and the other one based on images subject to a random uniform displacement of at most 4 pixels in (positive and negative) x and y direction to study the impact of offsets on feature importance. The resulting RFs achieve a classification error of 4.2% and 9.6%, respectively.

Discussion

Fig. 5.5 shows the cumulative feature importance (per triplet of layer index, wavelet scale, and direction, averaged over all trees in the respective RF) in handwritten digit classification and in facial landmark detection. Table 5.2 shows the corresponding cumulative feature importance for each layer.

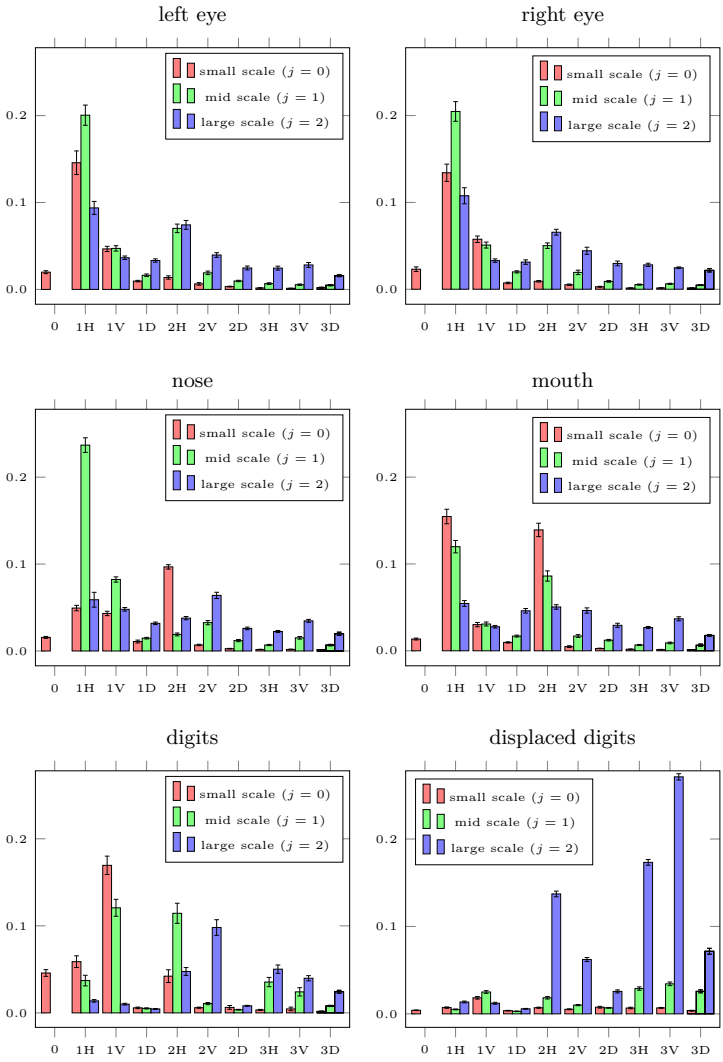

Fig. 5.5: Average cumulative feature importance and standard error for facial landmark detection and handwritten digit classification. The labels on the horizontal axis indicate the layer index $d \in \{0, 1, 2, 3\}$ and the wavelet direction (H: horizontal, V: vertical, D: diagonal).

For facial landmark detection, the features in layer 1 clearly have the highest importance, and the feature importance decreases with increasing layer index d. For handwritten digit classification using the unshifted MNIST images, the cumulative importance of the features in the second/third layer relative to those in the first layer is considerably higher than in facial landmark detection (see Table 5.2). For the translated MNIST images, the importance of the features in the second/third layer is significantly higher than those in the 0-th and in the first layer. An explanation for this observation could be as follows: In a classification task small sensitivity to translations is beneficial. Now, according to our theory (see Section 5.5.2) translation sensitivity, indeed, decreases with increasing layer index for average pooling as used here. For localization of landmarks, on the other hand, the RF needs features that are covariant on the fine grid of the input image thus favoring features in the layers closer to the root.

5.7. PROOFS

5.7.1. Proof of Lipschitz continuity of poolings

We verify the Lipschitz property $\|P(f) - P(h)\|_2 \leqslant R\|f - h\|_2$, for all $f, h \in H_N$, for the pooling operators in Section 5.2.3.

Sub-sampling

Pooling by sub-sampling is defined as

$$P : H_N \to H_{N/S}, \quad P(f)[n] = f[Sn], \quad n \in I_{N/S},$$

where $N/S \in \mathbb{N}$. Lipschitz continuity with $R = 1$ follows from

$$\begin{aligned} \|P(f) - P(h)\|_2^2 &= \sum_{n \in I_{N/S}} |f[Sn] - h[Sn]|^2 \\ &\leqslant \sum_{n \in I_N} |f[n] - h[n]|^2 = \|f - h\|_2^2, \quad \forall f, h \in H_N. \end{aligned}$$

Averaging

Pooling by averaging is defined as

$$P : H_N \to H_{N/S}, \quad P(f)[n] = \sum_{k=Sn}^{Sn+S-1} \alpha_{k-Sn} f[k],$$

for $n \in I_{N/S}$, where $N/S \in \mathbb{N}$. We start by setting $\alpha' := \max_{k\in\{0,\dots,S-1\}} |\alpha_k|$. Then,

$$\begin{aligned}
\|P(f) - P(h)\|_2^2 &= \sum_{n\in I_{N/S}} \Big| \sum_{k=Sn}^{Sn+S-1} \alpha_{k-Sn}(f[k] - h[k]) \Big|^2 \\
&\leqslant \sum_{n\in I_{N/S}} \Big| \sum_{k=Sn}^{Sn+S-1} \alpha' |f[k] - h[k]| \Big|^2 \\
&\leqslant \alpha'^2 S \sum_{n\in I_{N/S}} \sum_{k=Sn}^{Sn+S-1} \Big| f[k] - h[k] \Big|^2 \qquad (5.23) \\
&= \alpha'^2 S \sum_{n\in I_N} \Big| f[k] - h[k] \Big|^2 = \alpha'^2 S \|f - h\|_2^2,
\end{aligned}$$

where we used $\sum_{k\in I_S} |f[k] - h[k]| \leqslant S^{1/2}\|f - h\|_2$, $f, h \in H_S$, to get (5.23), see, e.g., (Golub and Van Loan, 2013, Equation 2.2.5).

Maximization

Pooling by maximization is defined as

$$P : H_N \to H_{N/S}, \quad P(f)[n] = \max_{k\in\{Sn,\dots,Sn+S-1\}} |f[k]|,$$

for $n \in I_{N/S}$, where $N/S \in \mathbb{N}$. We have

$$\begin{aligned}
\|P(f) - P(h)\|_2^2 = \sum_{n\in I_{N/S}} \Big| &\max_{k\in\{Sn,\dots,Sn+S-1\}} |f[k]| \\
&- \max_{k\in\{Sn,\dots,Sn+S-1\}} |h[k]| \Big|^2
\end{aligned}$$

$$\leqslant \sum_{n\in I_{N/S}} \max_{k\in\{Sn,\dots,Sn+S-1\}} \big|f[k]-h[k]\big|^2 \qquad (5.24)$$

$$\leqslant \sum_{n\in I_{N/S}} \sum_{k=0}^{S-1} |f[Sn+k]-h[Sn+k]|^2 \qquad (5.25)$$

$$= \|f-h\|_2^2,$$

where we employed the reverse triangle inequality

$$\big|\|f\|_\infty - \|h\|_\infty\big| \leqslant \|f-h\|_\infty, \quad f,h\in H_S,$$

to get (5.24), and in (5.25) we used $\|f\|_\infty \leqslant \|f\|_2$, $f\in H_S$, see, e.g., (Golub and Van Loan, 2013, Equation 2.2.6).

5.7.2. Proof of Theorem 5

We start by proving i). The key idea of the proof is—similarly to the proof of Proposition 7 in Section 3.6.8—to employ telescoping series arguments. For ease of notation, we let $f_q := U[q]f$ and $h_q := U[q]h$, for $f,h\in H_{N_1}$, $q\in\Lambda^d$. With (5.11) we have

$$|||\Phi_\Omega(f)-\Phi_\Omega(h)|||^2 = \sum_{d=0}^{D-1} \underbrace{\sum_{q\in\Lambda^d} ||(f_q-h_q)*\chi_d||_2^2}_{=:a_d}.$$

The key step is then to show that a_d can be upper-bounded according to

$$a_d \leqslant b_d - b_{d+1}, \qquad d=0,\dots,D-1, \qquad (5.26)$$

with $b_d := \sum_{q\in\Lambda^d}\|f_q-h_q\|_2^2$, for $d=0,\dots,D$, and to note that

$$\sum_{d=0}^{D-1} a_d \leqslant \sum_{d=0}^{D-1}(b_d-b_{d+1}) = b_0 - \underbrace{b_D}_{\geqslant 0} \leqslant b_0$$

$$= \sum_{q\in\Lambda^0}\|f_q-h_q\|_2^2 = \|f-h\|_2^2,$$

which then yields (5.10). Writing out (5.26), it follows that we need to establish

$$\sum_{q\in\Lambda^d} \|(f_q - h_q) * \chi_d\|_2^2 \leqslant \sum_{q\in\Lambda^d} ||f_q - h_q||_2^2 - \sum_{q\in\Lambda^{d+1}} \|f_q - h_q\|_2^2, \quad d = 0,\dots,D-1. \tag{5.27}$$

We start by examining the second sum on the RHS in (5.27). Every path

$$\tilde{q} \in \Lambda^{d+1} = \underbrace{\Lambda_1 \times \cdots \times \Lambda_d}_{=\Lambda^d} \times \Lambda_{d+1}$$

of length $d+1$ can be decomposed into a path $q \in \Lambda^d$ of length d and an index $\lambda_{d+1} \in \Lambda_{d+1}$ according to $\tilde{q} = (q, \lambda_{d+1})$. Thanks to (5.5) we have $U[\tilde{q}] = U[(q, \lambda_{d+1})] = U_{d+1}[\lambda_{d+1}]U[q]$, which yields

$$\sum_{\tilde{q}\in\Lambda^{d+1}} \|f_{\tilde{q}} - h_{\tilde{q}}\|_2^2 = \sum_{q\in\Lambda^d} \sum_{\lambda_{d+1}\in\Lambda_{d+1}} \|U_{d+1}[\lambda_{d+1}]f_q - U_{d+1}[\lambda_{d+1}]h_q\|_2^2. \tag{5.28}$$

Substituting (5.28) into (5.27) and rearranging terms, we obtain

$$\sum_{q\in\Lambda^d} \Big(\|(f_q - h_q) * \chi_d\|_2^2 \tag{5.29}$$

$$+ \sum_{\lambda_{d+1}\in\Lambda_{d+1}} \|U_{d+1}[\lambda_{d+1}]f_q - U_{d+1}[\lambda_{d+1}]h_q\|_2^2 \Big) \tag{5.30}$$

$$\leqslant \sum_{q\in\Lambda^d} ||f_q - h_q||_2^2, \quad d = 0,\dots,D-1. \tag{5.31}$$

We next note that the sum over the index set Λ_{d+1} inside the brackets in (5.29)-(5.30) satisfies

$$\sum_{\lambda_{d+1}\in\Lambda_{d+1}} \|U_{d+1}[\lambda_{d+1}]f_q - U_{d+1}[\lambda_{d+1}]h_q\|_2^2$$

$$= \sum_{\lambda_{d+1}\in\Lambda_{d+1}} \|P_{d+1}\big(\rho_{d+1}(f_q * g_{\lambda_{d+1}})\big) - P_{d+1}\big(\rho_{d+1}(h_q * g_{\lambda_{d+1}})\big)\|_2^2$$

$$\leqslant R_{d+1}^2 \sum_{\lambda_{d+1}\in\Lambda_{d+1}} \|\rho_{d+1}(f_q * g_{\lambda_{d+1}}) - \rho_{d+1}(h_q * g_{\lambda_{d+1}})\|_2^2 \qquad (5.32)$$

$$\leqslant R_{d+1}^2 L_{d+1}^2 \sum_{\lambda_{d+1}\in\Lambda_{d+1}} \|(f_q - h_q) * g_{\lambda_{d+1}}\|_2^2, \qquad (5.33)$$

where we employed the Lipschitz continuity of P_{d+1} in (5.32) and the Lipschitz continuity of ρ_{d+1} in (5.33). Substituting the sum over the index set Λ_{d+1} inside the brackets in (5.29)-(5.30) by the upper bound (5.33) yields

$$\sum_{q\in\Lambda^d} \Big(\|(f_q - h_q) * \chi_d\|_2^2$$

$$+ \sum_{\lambda_{d+1}\in\Lambda_{d+1}} \|U_{d+1}[\lambda_{d+1}]f_q - U_{d+1}[\lambda_{d+1}]h_q\|_2^2 \Big)$$

$$\leqslant \sum_{q\in\Lambda^d} \max\{1, R_{d+1}^2 L_{d+1}^2\} \Big(\|(f_q - h_q) * \chi_d\|_2^2 \qquad (5.34)$$

$$+ \sum_{\lambda_{d+1}\in\Lambda_{d+1}} \|(f_q - h_q) * g_{\lambda_{d+1}}\|_2^2 \Big), \qquad (5.35)$$

for $d = 0, \ldots, D-1$. As $\{g_{\lambda_{d+1}}\}_{\lambda_{d+1}\in\Lambda_{d+1}} \cup \{\chi_d\}$ are atoms of the convolutional set Ψ_{d+1}, and $f_q, h_q \in H_{N_{d+1}}$, we have

$$\|(f_q - h_q) * \chi_d\|_2^2 + \sum_{\lambda_{d+1}\in\Lambda_{d+1}} \|(f_q - h_q) * g_{\lambda_{d+1}}\|_2^2$$

$$\leqslant B_{d+1} \|f_q - h_q\|_2^2,$$

which, when used in (5.34)-(5.35) yields

$$\sum_{q\in\Lambda^d} \Big(\|(f_q - h_q) * \chi_d\|_2^2$$

$$+ \sum_{\lambda_{d+1}\in\Lambda_{d+1}} \|U_{d+1}[\lambda_{d+1}]f_q - U_{d+1}[\lambda_{d+1}]h_q\|_2^2 \Big)$$

$$\leqslant \sum_{q\in\Lambda^d} \max\{B_{d+1}, B_{d+1} R_{d+1}^2 L_{d+1}^2\} \|f_q - h_q\|_2^2, \qquad (5.36)$$

for $d = 0, \ldots, D-1$. Finally, invoking (5.9) in (5.36) we get (5.29)-(5.31) and hence (5.26). This completes the proof of i).

We continue with ii). The key step in establishing (5.12) is to show that for $\rho_d(0) = 0$ and $P_d(0) = 0$, for $d \in \{1, \ldots, D-1\}$, the feature extractor Φ_Ω satisfies $\Phi_\Omega(0) = 0$, and to employ (5.10) with $h = 0$ which yields $|||\Phi(f)||| \leqslant \|f\|$, for $f \in H_{N_1}$. It remains to prove that $\Phi_\Omega(h) = 0$ for $h = 0$. For $h = 0$, the operator U_d, $d \in \{1, 2, \ldots, D\}$, defined in (5.4) satisfies

$$(U_d[\lambda_d]h) = P_d\big(\underbrace{\rho_d\big(\underbrace{\underbrace{(h * g_{\lambda_d})}_{=0}}_{=0}\big)}_{=0}\big),$$

for $\lambda_d \in \Lambda_d$, by assumption. With the definition of $U[q]$ in (5.5) this then yields $(U[q]h) = 0$ for $h = 0$ and all $q \in \Lambda^d$. $\Phi_\Omega(0) = 0$ finally follows from

$$\Phi_\Omega(h) = \bigcup_{d=0}^{D-1} \big\{ \underbrace{\big(U[q]h\big) * \chi_d}_{=0} \big\}_{q \in \Lambda^d} = 0. \tag{5.37}$$

We proceed to iii). The proof of the deformation sensitivity bound (5.14) is based on two key ingredients. The first one is the Lipschitz continuity result stated in (5.10). The second ingredient, stated in Proposition 10 in Section 5.7.3, is an upper bound on the deformation error $\|f - F_\tau f\|_2$ given by

$$\|f - F_\tau f\|_2 \leqslant 4K N_1^{1/2} \|\tau\|_\infty^{1/2}, \tag{5.38}$$

where $f \in C_{\mathrm{CART}}^{N_1, K}$. We now show how (5.10) and (5.38) can be combined to establish (5.14). To this end, we first apply (5.10) with $h := (F_\tau f)$ to get

$$|||\Phi_\Omega(f) - \Phi_\Omega(F_\tau f)||| \leqslant \|f - F_\tau f\|_2, \tag{5.39}$$

for $f \in C_{\mathrm{CART}}^{N_1, K} \subseteq H_{N_1}$, $N_1 \in \mathbb{N}$, and $K > 0$, and then replace the RHS of (5.39) by the RHS of (5.38). This completes the proof of iii).

5.7.3. Proof of Proposition 10

Proposition 10. *For every $N \in \mathbb{N}$, every $K > 0$, and every $\tau : \mathbb{R} \to [-1, 1]$, we have*

$$\|f - F_\tau f\|_2 \leq 4KN^{1/2}\|\tau\|_\infty^{1/2}, \tag{5.40}$$

for all $f \in \mathcal{C}_{\mathrm{CART}}^{N,K}$.

Remark 11. *As already mentioned at the end of Section 5.4, excluding the interval boundary points a, b in the definition of sampled cartoon functions $\mathcal{C}_{\mathrm{CART}}^{N,K}$ (see Definition 11 in Section 5.4) is necessary for technical reasons. Specifically, without imposing this exclusion, we can not expect to get deformation sensitivity results of the form* (5.40)*. This can be seen as follows. Let us assume that we seek a bound of the form $\|f - F_\tau f\|_2 \leqslant C_{N,K}\|\tau\|_\infty^\alpha$, for some $C_{N,K} > 0$ and some $\alpha > 0$, that applies to all $f[n] = c(n/N)$, $n \in I_N$, with $c \in \mathcal{C}_{\mathrm{CART}}^K$. Take $\tau(x) = 1/N$, in which case the deformation $(F_\tau f)[n] = c(n/N - 1/N)$ amounts to a simple translation by $1/N$ and $\|\tau\|_\infty = 1/N \leqslant 1$. Let $c(x) = \mathbb{1}_{[0,2/N]}(x)$. Then $c \in \mathcal{C}_{\mathrm{CART}}^K$ for $K = 1$ and $\|f - F_\tau f\|_2 = \sqrt{2}$, which obviously does not decay with $\|\tau\|_\infty^\alpha = N^{-\alpha}$ for some $\alpha > 0$. We note that this phenomenon occurs only in the discrete case.*

Proof. The proof of (5.40) is based on judiciously combining deformation sensitivity bounds for the sampled components $c_1(n/N), c_2(n/N)$, $n \in I_N$, in $(c_1 + \mathbb{1}_{[a,b]}c_2) \in \mathcal{C}_{\mathrm{CART}}^K$, and the sampled indicator function $\mathbb{1}_{[a,b]}(n/N)$, $n \in I_N$. The first bound, stated in Lemma 10 below, reads

$$\|f - F_\tau f\|_2 \leq CN^{1/2}\|\tau\|_\infty, \tag{5.41}$$

and applies to discrete-time signals $f[n] = f(n/N)$, $n \in I_N$, with $f : \mathbb{R} \to \mathbb{C}$ satisfying the Lipschitz property with Lipschitz constant C. The second bound we need, stated in Lemma 11 below, is given by

$$\|\mathbb{1}_{[a,b]}^N - F_\tau \mathbb{1}_{[a,b]}^N\|_2 \leqslant 2N^{1/2}\|\tau\|_\infty^{1/2}, \tag{5.42}$$

and applies to sampled indicator functions $\mathbb{1}_{[a,b]}^N[n] := \mathbb{1}_{[a,b]}(n/N)$, $n \in I_N$, with $a, b \notin \{0, \frac{1}{N}, \ldots, \frac{N-1}{N}\}$. We now show how (5.41) and

(5.42) can be combined to establish (5.40). For a sampled cartoon function $f \in \mathcal{C}^{N,K}_{\mathrm{CART}}$, i.e.,

$$f[n] = c_1(n/N) + \mathbb{1}_{[a,b]}(n/N)c_2(n/N) =: f_1[n] + \mathbb{1}^N_{[a,b]}[n]f_2[n],$$

where $n \in I_N$, we have

$$\begin{aligned}\|f - F_\tau f\|_2 &\leqslant \|f_1 - F_\tau f_1\|_2 + \|\mathbb{1}^N_{[a,b]}(f_2 - F_\tau f_2)\|_2 \\ &+ \|(\mathbb{1}^N_{[a,b]} - F_\tau \mathbb{1}^N_{[a,b]})(F_\tau f_2)\|_2 \qquad (5.43)\\ &\leqslant \|f_1 - F_\tau f_1\|_2 + \|f_2 - F_\tau f_2\|_2 + \|\mathbb{1}^N_{[a,b]} - F_\tau \mathbb{1}^N_{[a,b]}\|_2 \|F_\tau f_2\|_\infty,\end{aligned}$$

where in (5.43) we used

$$\begin{aligned}\big(F_\tau(\mathbb{1}^N_{[a,b]}f_2)\big)[n] &= (\mathbb{1}_{[a,b]}c_2)(n/N - \tau(n/N)) \\ &= \mathbb{1}_{[a,b]}(n/N - \tau(n/N))c_2((n/N - \tau(n/N))) \\ &= (F_\tau \mathbb{1}^N_{[a,b]})[n](F_\tau f_2)[n].\end{aligned}$$

With the upper bounds (5.41) and (5.42), invoking properties of $\mathcal{C}^{N,K}_{\mathrm{CART}}$ (namely, (i) c_1, c_2 satisfy the Lipschitz property with Lipschitz constant $C = K$ and hence $f_1[n] = c_1(n/N), f_2[n] = c_2(n/N), n \in I_N$, satisfy (5.41) with $C = K$, and (ii) $\|F_\tau f_2\|_\infty = \sup_{n\in I_N} |(F_\tau f_2)[n]| = \sup_{n\in I_N} |c_2(n/N - \tau(n/N))| \leqslant \sup_{x\in\mathbb{R}} |c_2(x)| = \|c_2\|_\infty \leqslant K$), this yields

$$\begin{aligned}\|f - F_\tau f\|_2 &\leqslant 2KN^{1/2}\,\|\tau\|_\infty + 2KN^{1/2}\|\tau\|^{1/2}_\infty \\ &\leqslant 4KN^{1/2}\|\tau\|^{1/2}_\infty,\end{aligned}$$

where in the last step we used $\|\tau\|_\infty \leqslant \|\tau\|^{1/2}_\infty$, which is thanks to the assumption $\|\tau\|_\infty \leqslant 1$. This completes the proof of (5.40). □

It remains to establish (5.41) and (5.42).

Lemma 10. *Let $c : \mathbb{R} \to \mathbb{C}$ be Lipschitz-continuous with Lipschitz constant C. Let further $f[n] := c(n/N)$, $n \in I_N$. Then,*

$$\|f - F_\tau f\|_2 \leq CN^{1/2}\|\tau\|_\infty.$$

Proof. Invoking the Lipschitz property of c according to

$$\begin{aligned}\|f - F_\tau f\|_2^2 &= \sum_{n\in I_N} |f[n] - (F_\tau f)[n]|^2 \\ &= \sum_{n\in I_N} |c(n/N) - c(n/N - \tau(n/N))|^2 \\ &\leqslant C^2 \sum_{n\in I_N} |\tau(n/N)|^2 \leqslant C^2 N \|\tau\|_\infty^2\end{aligned}$$

completes the proof. □

We continue with a deformation sensitivity result for sampled indicator functions $\mathbb{1}_{[a,b]}(x)$.

Lemma 11. *Let $[a,b] \subseteq [0,1]$ and set $\mathbb{1}^N_{[a,b]}[n] := \mathbb{1}_{[a,b]}(n/N)$, $n \in I_N$, with $a, b \notin \{0, \frac{1}{N}, \ldots, \frac{N-1}{N}\}$. Then, we have*

$$\|\mathbb{1}^N_{[a,b]} - F_\tau \mathbb{1}^N_{[a,b]}\|_2 \leq 2N^{1/2} \|\tau\|_\infty^{1/2}.$$

Proof. In order to upper-bound

$$\begin{aligned}\|\mathbb{1}^N_{[a,b]} - F_\tau \mathbb{1}^N_{[a,b]}\|_2^2 &= \sum_{n\in I_N} |\mathbb{1}^N_{[a,b]}[n] - (F_\tau \mathbb{1}^N_{[a,b]})[n]|^2 \\ &= \sum_{n\in I_N} |\mathbb{1}_{[a,b]}(n/N) - \mathbb{1}_{[a,b]}(n/N - \tau(n/N))|^2,\end{aligned}$$

we first note that the summand $h(n) := |\mathbb{1}_{[a,b]}(n/N) - \mathbb{1}_{[a,b]}(n/N - \tau(n/N))|^2$ satisfies $h(n) = 1$, for $n \in S$, where

$$\begin{aligned}S := &\Big\{ n \in I_N \,\Big|\, \frac{n}{N} \in [a,b] \;\text{ and }\; \frac{n}{N} - \tau\Big(\frac{n}{N}\Big) \notin [a,b] \Big\} \\ \cup &\Big\{ n \in I_N \,\Big|\, \frac{n}{N} \notin [a,b] \;\text{ and }\; \frac{n}{N} - \tau\Big(\frac{n}{N}\Big) \in [a,b] \Big\},\end{aligned}$$

and $h(n) = 0$, for $n \in I_N \backslash S$. Thanks to $a, b \notin \{0, \frac{1}{N}, \ldots, \frac{N-1}{N}\}$, we have $S \subseteq \Sigma$, where

$$\Sigma := \Big\{ n \in \mathbb{Z} \,\Big|\, \Big|\frac{n}{N} - a\Big| < \|\tau\|_\infty \Big\} \cup \Big\{ n \in \mathbb{Z} \,\Big|\, \Big|\frac{n}{N} - b\Big| < \|\tau\|_\infty \Big\}.$$

The cardinality of the set Σ can be upper-bounded by $2\,\frac{2\|\tau\|_\infty}{1/N}$, which then yields

$$\begin{aligned}\|\mathbb{1}^N_{[a,b]} - F_\tau \mathbb{1}^N_{[a,b]}\|_2^2 &= \sum_{n\in I_N} |h(n)|^2 = \sum_{n\in S} 1 \\ &\leqslant \sum_{n\in\Sigma} 1 \leqslant 4N\|\tau\|_\infty. \end{aligned} \tag{5.44}$$

This completes the proof. □

Remark 12. *For general $a, b \in [0,1]$, i.e., when we drop the assumption $a, b \notin \{0, \frac{1}{N}, \ldots, \frac{N-1}{N}\}$, it follows that $S \subseteq \Sigma'$, where*

$$\Sigma' := \left\{n \in \mathbb{Z} \,\middle|\, \left|\frac{n}{N} - a\right| \leqslant \|\tau\|_\infty\right\} \cup \left\{n \in \mathbb{Z} \,\middle|\, \left|\frac{n}{N} - b\right| \leqslant \|\tau\|_\infty\right\}.$$

Noting that the cardinality of Σ' can be upper-bounded by $2\left(\frac{2\|\tau\|_\infty}{1/N} + 1\right) = 4N\|\tau\|_\infty + 2$, this then yields (similarly to (5.44))

$$\|\mathbb{1}^N_{[a,b]} - F_\tau \mathbb{1}^N_{[a,b]}\|_2^2 \leqslant \sum_{n\in\Sigma} 1 \leqslant 4N\|\tau\|_\infty + 2,$$

which shows that the deformation error—for general $a, b \in [0,1]$—does not decay with $\|\tau\|_\infty^\alpha$ for some $\alpha > 0$ (see also the example in Remark 11).

5.7.4. Proof of Theorem 6

We start by establishing i). For ease of notation, again, we let $f_q := U[q]f$ and $h_q := U[q]h$, for $f, h \in H_{N_1}$, $q \in \Lambda^d$. We have

$$|||\Phi^d_\Omega(f) - \Phi^d_\Omega(h)|||^2 = \sum_{q\in\Lambda^d} ||(f_q - h_q) * \chi_d||_2^2 \tag{5.45}$$

$$\leqslant \|\chi_d\|_1^2 \underbrace{\sum_{q\in\Lambda^d} ||(f_q - h_q)||_2^2}_{=:a_d}, \tag{5.46}$$

where (5.46) follows by Young's inequality (Folland, 2015, Proposition 2.3.9).

Remark 13. *We emphasize that* (5.45) *can also be upper-bounded by* $B_{d+1}\sum_{q\in\Lambda^d}||(f_q - h_q)||_2^2$, *which follows from the fact that* $\{g_{\lambda_{d+1}}\}_{\lambda_{d+1}\in\Lambda_{d+1}} \cup \{\chi_d\}$ *are atoms of the convolutional set* Ψ_{d+1} *with Bessel bound* B_{d+1}. *Hence, one can substitute* $\|\chi_d\|_1$ *in* (5.17) *by* $\sqrt{B_{d+1}}$.

The key step is then to show that a_d can be upper-bounded according to

$$a_k \leqslant (B_k L_k^2 R_k^2) a_{k-1}, \qquad k = 1, \ldots, d, \tag{5.47}$$

and to note that

$$\begin{aligned} a_d &\leqslant (B_d L_d^2 R_d^2) a_{d-1} \leqslant \ldots \leqslant \Big(\prod_{k=1}^{d} B_k L_k^2 R_k^2\Big) a_0 \\ &= \Big(\prod_{k=1}^{d} B_k L_k^2 R_k^2\Big) \sum_{q\in\Lambda_1^0} \|f_q - h_q\|_2^2 \\ &= \Big(\prod_{k=1}^{d} B_k L_k^2 R_k^2\Big) \|f - h\|_2^2, \end{aligned}$$

which yields (5.18). We now establish (5.47). Every path

$$\tilde{q} \in \Lambda_1^k = \underbrace{\Lambda_1 \times \cdots \times \Lambda_{k-1}}_{=\Lambda^{k-1}} \times \Lambda_k$$

of length k can be decomposed into a path $q \in \Lambda^{k-1}$ of length $k-1$ and an index $\lambda_k \in \Lambda_k$ according to $\tilde{q} = (q, \lambda_k)$. Thanks to (5.5) we have $U[\tilde{q}] = U[(q, \lambda_k)] = U_k[\lambda_k]U[q]$, which yields

$$\sum_{\tilde{q}\in\Lambda^k} \|f_{\tilde{q}} - h_{\tilde{q}}\|_2^2 = \sum_{q\in\Lambda^{k-1}} \sum_{\lambda_k\in\Lambda_k} \|U_k[\lambda_k] f_q - U_k[\lambda_k] h_q\|_2^2. \tag{5.48}$$

We next note that the term inside the sums on the RHS in (5.48) satisfies

$$\begin{aligned} \|U_k[\lambda_k] f_q - U_k[\lambda_k] h_q\|_2^2 &= \|P_k\big(\rho_k(f_q * g_{\lambda_k})\big) - P_k\big(\rho_k(h_q * g_{\lambda_k})\big)\|_2^2 \\ &\leqslant L_k^2 R_k^2 \|(f_q - h_q) * g_{\lambda_k}\|_2^2, \end{aligned} \tag{5.49}$$

where we used the Lipschitz continuity of P_k and ρ_k with Lipschitz constants $R_k > 0$ and $L_k > 0$, respectively. As $\{g_{\lambda_k}\}_{\lambda_k \in \Lambda_k} \cup \{\chi_{k-1}\}$ are the atoms of the convolutional set Ψ_k, and $f_q, h_q \in H_{N_k}$ by (5.5), we have

$$\sum_{\lambda_k \in \Lambda_k} \|(f_q - h_q) * g_{\lambda_k}\|_2^2 \leqslant B_k \|f_q - h_q\|_2^2,$$

which, when used in (5.48) together with (5.49), yields

$$\sum_{\tilde{q} \in \Lambda^k} \|f_{\tilde{q}} - h_{\tilde{q}}\|_2^2 \leqslant B_k L_k^2 R_k^2 \sum_{q \in \Lambda^{k-1}} \|f_q - h_q\|_2^2,$$

and hence establishes (5.47), thereby completing the proof of i).

We now turn to ii). The proof of (5.19) follows—as in the proof of ii) in Theorem 5 in Section 5.7.2—from (5.18) together with $\Phi_\Omega^d(h) = \{(U[q]h) * \chi_d\}_{q \in \Lambda^d} = 0$ for $h = 0$, see (5.37).

We continue with iii). The proof of the deformation sensitivity bound (5.20) is based on two key ingredients. The first one is the Lipschitz continuity result in (5.18). The second ingredient is, again, the deformation sensitivity bound (5.40) stated in Proposition 10 in Section 5.7.3. Combining (5.18) and (5.40)—as in the proof of iii) in Theorem 5 in Section 5.7.2—then establishes (5.20) and completes the proof of iii).

We proceed to iv). For ease of notation, again, we let $f_q := U[q]f$, for $f \in H_{N_1}$, $q \in \Lambda^d$. Thanks to (5.5), we have $f_q \in H_{N_{d+1}}$, for $q \in \Lambda^d$. The key step in establishing (5.21) is to show that the operator U_k, $k \in \{1, 2, \ldots, d\}$, defined in (5.4) satisfies the relation

$$(U_k[\lambda_k] T_m f) = T_{m/S_k}(U_k[\lambda_k] f), \tag{5.50}$$

for $f \in H_{N_k}$, $m \in \mathbb{Z}$ with $\frac{m}{S_k} \in \mathbb{Z}$, and $\lambda_k \in \Lambda_k$. With the definition of $U[q]$ in (5.5) this then yields

$$(U[q] T_m f) = T_{m/(S_1 \cdots S_d)}(U[q] f), \tag{5.51}$$

for $f \in H_{N_1}$, $m \in \mathbb{Z}$ with $\frac{m}{S_1 \ldots S_d} \in \mathbb{Z}$, and $q \in \Lambda^d$. The identity (5.21) is then a direct consequence of (5.51) and the translation-covariance of

the circular convolution operator (which holds thanks to $\frac{m}{S_1 \dots S_d} \in \mathbb{Z}$):

$$\begin{aligned}\Phi_\Omega^d(T_m f) &= \big\{\big(U[q]T_m f\big) * \chi_d\big\}_{q\in\Lambda^d} = \big\{\big(T_{m/(S_1\cdots S_d)}U[q]f\big) * \chi_d\big\}_{q\in\Lambda^d} \\ &= \big\{T_{m/(S_1\cdots S_d)}\big((U[q]f) * \chi_d\big)\big\}_{q\in\Lambda^d} = T_{m/(S_1\cdots S_d)}\Phi_\Omega^d(f),\end{aligned}$$

for $f \in H_{N_1}$ and $m \in \mathbb{Z}$ with $\frac{m}{S_1 \dots S_d} \in \mathbb{Z}$. It remains to establish (5.50):

$$\begin{aligned}(U_k[\lambda_k]T_m f) &= \Big(P_k\big(\rho_k((T_m f) * g_{\lambda_k})\big)\Big) \\ &= \Big(P_k\big(\rho_k(T_m(f * g_{\lambda_k}))\big)\Big) \qquad (5.52) \\ &= \Big(P_k\big(T_m(\rho_k(f * g_{\lambda_k}))\big)\Big), \qquad (5.53)\end{aligned}$$

where in (5.52) we used the translation covariance of the circular convolution operator (which holds thanks to $m \in \mathbb{Z}$), and in (5.53) we used the fact that point-wise non-linearities commute with the translation operator thanks to

$$(\rho_k T_m f)[n] = \rho_k((T_m f)[n]) = \rho_k(f[n-m]) = (T_m \rho_k f)[n],$$

for $f \in H_{N_k}$, $n \in I_{N_k}$, and $m \in \mathbb{Z}$. Next, we note that the pooling operators P_k in Section 5.2.3 (namely, sub-sampling, average pooling, and max-pooling) can all be written as $(P_k f)[n] = (P'_k f)[S_k n]$, for some P'_k that commutes with the translation operator, namely, for (i) sub-sampling $(P'_k f)[n] = f[n]$, with $(P'_k T_m f)[n] = (T_m f)[n] = f[n-m] = (T_m P'_k f)[n]$, (ii) average pooling $(P'_k f)[n] = \sum_{l=n}^{n+S_k-1} \alpha_{l-n} f[l]$ with

$$\begin{aligned}(P'_k T_m f)[n] &= \sum_{l=n}^{n+S_k-1} \alpha_{l-n} f[l-m] = \sum_{l'=(n-m)}^{(n-m)+S_k-1} \alpha_{l-(n-m)} f[l'] \\ &= (T_m P'_k f)[n],\end{aligned}$$

and for (iii) max-pooling $(P'_k f)[n] = \max_{l\in\{n,\dots,n+S_k-1\}} |f[l]|$ with

$$(P'_k T_m f)[n] = \max_{l\in\{n,\dots,n+S_k-1\}} |f[l-m]|$$

$$
\begin{aligned}
&= \max_{(l-m)\in\{n-m,\ldots,(n-m)+S_k-1\}} |f[l-m]| \\
&= \max_{l'\in\{(n-m),\ldots,(n-m)+S_k-1\}} |f[l']| \\
&= (T_m P_k' f)[n],
\end{aligned}
$$

in all three cases for $f \in H_{N_k}$, $n \in I_{N_k}$, and $m \in \mathbb{Z}$. This then yields

$$
\begin{aligned}
(P_k T_m f)[n] &= (P_k' T_m f)[S_k n] = (T_m P_k' f)[S_k n] \\
&= P_k'(f)[S_k n - m] = P_k'(f)[S_k(n - S_k^{-1} m)] \\
&= P_k(f)[n - S_k^{-1} m] = (T_{m/S_k} P_k f)[n], \qquad (5.54)
\end{aligned}
$$

for $f \in H_{N_k}$ and $n \in I_{N_{k+1}}$. Here, we used $m/S_k \in \mathbb{Z}$, which is by assumption. Substituting (5.54) into (5.53) finally yields

$$
(U_k[\lambda_k] T_m f) = T_{m/S_k} U_k[\lambda_k] f,
$$

for $f \in H_{N_k}$, $m \in \mathbb{Z}$ with $\frac{m}{S_k} \in \mathbb{Z}$, and $\lambda_k \in \Lambda_k$. This completes the proof of (5.50) and hence establishes (5.21).

References

Adams, A. R. (1975), *Sobolev spaces*, Academic Press, Amsterdam.

Ali, S. T., Antoine, J. P., and Gazeau, J. P. (1993), "Continuous frames in Hilbert spaces," *Annals of Physics*, vol. 222, no. 1, pp. 1–37.

Andén, J. and Mallat, S. (2014), "Deep scattering spectrum," *IEEE Trans. Signal Process.*, vol. 62, no. 16, pp. 4114–4128.

Angelova, A., Abu-Mostafa, Y., and Perona, P. (2005), "Pruning training sets for learning of object categories," in *Proc. of IEEE Conf. Comp. Vision Pattern Recog. (CVPR)*, pp. 494–501.

Antoine, J. P., Murrenzi, R., Vandergheynst, P., and Ali, S. T. (2008), *Two-dimensional wavelets and their relatives*, Cambridge University Press, Cambridge.

Arivazhagan, S., Ganesan, L., and Kumar, T. S. (2006), "Texture classification using ridgelet transform," *Pattern Recognition Letters*, vol. 27, no. 16, pp. 1875–1883.

Bengio, Y., Courville, A., and Vincent, P. (2013), "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828.

Bishop, C. M. (2009), *Pattern recognition and machine learning*, Springer, New York, NY.

Bölcskei, H. and Hlawatsch, F. (1997), "Discrete Zak transforms, polyphase transforms, and applications," *IEEE Trans. Sig. Process.*, vol. 45, no. 4, pp. 851–866.

Bölcskei, H., Hlawatsch, F., and Feichtinger, H. G. (1998), "Frame-theoretic analysis of oversampled filter banks," *IEEE Trans. Signal Process.*, vol. 46, no. 12, pp. 3256–3268.

Breiman, L. (2001), "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32.

Breiman, L., Friedman, J., Stone, C. J., and Olshen, R. A. (1984), *Classification and regression trees*, CRC Press, Boca Raton, FL.

Brent, R. P., Osborn, J. H., and Smith, W. D. (2015), "Note on best possible bounds for determinants of matrices close to the identity matrix," *Linear Algebra and its Applications*, vol. 466, pp. 21–26.

Bruna, J. and Mallat, S. (2013), "Invariant scattering convolution networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1872–1886.

Candès, E. J. (1998), *Ridgelets: Theory and applications*, Ph.D. thesis, Stanford University.

Candès, E. J., Demanet, L., Donoho, D., and Ying, L. (2006), "Fast discrete curvelet transforms," *Multiscale Modeling and Simulation*, vol. 5, no. 3, pp. 861–899.

Candès, E. J. and Donoho, D. L. (1999), "Ridgelets: A key to higher-dimensional intermittency?" *Philos. Trans. R. Soc. London Ser. A*, vol. 357, no. 1760, pp. 2495–2509.

——— (2004), "New tight frames of curvelets and optimal representations of objects with piecewise C^2 singularities," *Comm. Pure Appl. Math.*, vol. 57, no. 2, pp. 219–266.

——— (2005), "Continuous curvelet transform: II. Discretization and frames," *Appl. Comput. Harmon. Anal.*, vol. 19, no. 2, pp. 198–222.

Canny, J. (1986), "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 6, pp. 679–698.

do Carmo, M. P. (2013), *Riemannian geometry*, Birkhäuser, Basel, 3rd ed.

Chen, G. Y., Bui, T. D., and Krżyzak, A. (2005), "Rotation invariant pattern recognition using ridgelets, wavelet cycle-spinning and Fourier features," *Pattern Recognition*, vol. 38, no. 12, pp. 2314–2322.

Chen, S., Cowan, C., and Grant, P. M. (1991), "Orthogonal least squares learning algorithm for radial basis function networks," *IEEE Trans. Neural Netw.*, vol. 2, no. 2, pp. 302–309.

Christensen, O. (2003), *An introduction to frames and Riesz bases*, Springer, Basel.

Comenetz, M. (2002), *Calculus: The elements*, World Scientific, New Jersey, NJ.

Cortes, C. and Vapnik, V. (1995), "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297.

Czaja, W. and Li, W. (2017), "Analysis of time-frequency scattering transforms," *Appl. Comput. Harmon. Anal., to appear.*

Dantone, M., Gall, J., Fanelli, G., and Van Gool, L. (2012), "Real-time facial feature detection using conditional regression forests," in *Proc. of IEEE Conf. Comp. Vision Pattern Recog. (CVPR)*, pp. 2578–2585.

Daubechies, I. (1992), *Ten lectures on wavelets*, Society for Industrial and Applied Mathematics, Philadelphia, PA.

Daubechies, I., Grossmann, A., and Meyer, Y. (1986), "Painless non-orthogonal expansions," *J. Math. Phys.*, vol. 27, no. 5, pp. 1271–1283.

Daubechies, I., Landau, H. J., and Landau, Z. (1995), "Gabor time-frequency lattices and the Wexler-Raz identity," *J. Fourier Anal. Appl.*, vol. 1, no. 4, pp. 438–478.

Davis, S. and Mermelstein, P. (1980), "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," *IEEE Trans. Acoust., Speech, and Signal Process.*, vol. 28, no. 4, pp. 357–366.

Dettori, L. and Semler, L. (2007), "A comparison of wavelet, ridgelet, and curvelet-based texture classification algorithms in computed tomography," *Computers in Biology and Medicine*, vol. 37, no. 4, pp. 486–498.

DiBenedetto, E. (2002), *Real analysis*, Birkhäuser, New York, NY.

Donoho, D. L. (2001), "Sparse components of images and optimal atomic decompositions," *Constructive Approximation*, vol. 17, no. 3, pp. 353–382.

Duda, R. O., Hart, P. E., and Stork, D. G. (2001), *Pattern classification*, John Wiley, New York, NY, 2nd ed.

Ellis, D., Zeng, Z., and McDermott, J. (2011), "Classifying soundtracks with audio texture features," in *Proc. of IEEE International Conference on Acoust., Speech, and Signal Process. (ICASSP)*, pp. 5880–5883.

Federer, H. (1969), *Geometric measure theory*, Springer, Berlin, 1st ed.

Folland, G. B. (2015), *A course in abstract harmonic analysis*, vol. 29, CRC Press, Boca Raton, FL.

Frazier, M., Jawerth, B., and Weiss, G. (1991), *Littlewood-Paley theory and the study of function spaces*, American Mathematical Society, Providence, RI.

Glorot, X. and Bengio, Y. (2010), "Understanding the difficulty of training deep feedforward neural networks," in *Proc. of International Conference on Artificial Intelligence and Statistics (AISTATS)*, pp. 249–256.

Glorot, X., Bordes, A., and Bengio, Y. (2011), "Deep sparse rectifier neural networks," in *Proc. of International Conference on Artificial Intelligence and Statistics (AISTATS)*, pp. 315–323.

Golub, G. H. and Van Loan, C. F. (2013), *Matrix computations*, Johns Hopkins University Press, Baltimore, MD.

Goodfellow, I., Bengio, Y., and Courville, A. (2016), *Deep Learning*, MIT Press, Cambridge, MA. `http://www.deeplearningbook.org`.

Grafakos, L. (2008), *Classical Fourier analysis*, Springer, New York, NY, 2nd ed.

——— (2009), *Modern Fourier analysis*, Springer, New York, NY, 2nd ed.

Gray, A. (2004), *Tubes*, Springer, Boston, MA, 2nd ed.

Griffin, G., Holub, A., and Perona, P. (2007), "Caltech-256 object category dataset," `http://authors.library.caltech.edu/7694/`.

Gröchenig, K. (2001), *Foundations of time-frequency analysis*, Birkhäuser, Boston, MA.

Gröchenig, K., Janssen, A. J. E. M., Kaiblinger, N., and Pfander, G. E. (2003), "Note on B-Splines, wavelet scaling functions, and Gabor frames," *IEEE Trans. Inf. Theory*, vol. 49, no. 12, pp. 3318–3320.

Gröchenig, K. and Samarah, S. (2000), "Nonlinear approximation with local Fourier bases," *Constr. Approx.*, vol. 16, no. 3, pp. 317–331.

Grohs, P. (2012), "Ridgelet-type frame decompositions for Sobolev spaces related to linear transport," *J. Fourier Anal. Appl.*, vol. 18, no. 2, pp. 309–325.

Grohs, P., Keiper, S., Kutyniok, G., and Schäfer, M. (2015), "Cartoon approximation with α-curvelets," *J. Fourier Anal. Appl.*, pp. 1–59.

Grohs, P. and Kutyniok, G. (2014), "Parabolic molecules," *Foundations of Computational Mathematics*, vol. 14, no. 2, pp. 299–337.

Grohs, P., Wiatowski, T., and Bölcskei, H. (2016), "Deep convolutional neural networks on cartoon functions," in *Proc. of IEEE International Symposium on Information Theory (ISIT)*, pp. 1163–1167.

——— (2017), "Energy decay and conservation in deep convolutional neural networks," in *Proc. of IEEE International Symposium on Information Theory (ISIT)*, pp. 1356–1360.

Guo, K., Kutyniok, G., and Labate, D. (2006), "Sparse multidimensional representations using anisotropic dilation and shear operators," in G. Chen and M. J. Lai (Eds.), *Wavelets and Splines*, pp. 189–201, Nashboro Press, Brentwood, TN.

He, K., Zhang, X., Ren, S., and Sun, J. (2015), "Deep residual learning for image recognition," in *Proc. of IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778.

Holschneider, M., Kronland-Martinet, R., Morlet, J., and Tchamitchian, P. (1989), "A real-time algorithm for signal analysis with the help of the wavelet transform," in *Wavelets*, pp. 286–297.

Huang, F. J. and LeCun, Y. (2006), "Large-scale learning with SVM and convolutional nets for generic object categorization," in *Proc. of IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 284–291.

Janssen, A. J. E. M. (1995), "Duality and biorthogonality for Weyl-Heisenberg frames," *J. Fourier Anal. Appl.*, vol. 1, no. 4, pp. 403–436.

——— (1998), "The duality condition for Weyl-Heisenberg frames," in

H. G. Feichtinger and T. Strohmer (Eds.), *Gabor analysis: Theory and applications*, pp. 33–84, Birkhäuser, Boston, MA.

Jarrett, K., Kavukcuoglu, K., Ranzato, M., and LeCun, Y. (2009), "What is the best multi-stage architecture for object recognition?" in *Proc. of IEEE International Conference on Computer Vision (ICCV)*, pp. 2146–2153.

Kaiser, G. (1994), *A friendly guide to wavelets*, Birkhäuser, New York, NY.

Krizhevsky, A. (2009), *Learning multiple layers of features from tiny images*, Master's thesis, University of Toronto.

Krizhevsky, A., Sutskever, I., and Hinton, G. (2012), "Imagenet classification with deep convolutional neural networks," in *Proc. of Int. Conf. on Neural Information Processing Systems (NIPS)*, pp. 1097–1105.

Kutyniok, G. and Donoho, D. L. (2013), "Microlocal analysis of the geometric separation problem," *Comm. Pure Appl. Math.*, vol. 66, no. 1, pp. 1–47.

Kutyniok, G. and Labate, D. (2012a), "Introduction to shearlets," in Kutyniok and Labate (2012b), pp. 1–38.

Kutyniok, G. and Labate, D. (Eds.) (2012b), *Shearlets: Multiscale analysis for multivariate data*, Birkhäuser, Basel.

LeCun, Y., Bengio, Y., and Hinton, G. (2015), "Deep learning," *Nature*, vol. 521, pp. 436–444.

LeCun, Y., Boser, B., Denker, J., Henderson, D., Howard, R., Hubbard, W., and Jackel, L. (1990), "Handwritten digit recognition with a back-propagation network," in *Proc. of International Conference on Neural Information Processing Systems (NIPS)*, pp. 396–404.

LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998), "Gradient-based learning applied to document recognition," in *Proc. of the IEEE*, pp. 2278–2324.

LeCun, Y. and Cortes, C. (1998), "The MNIST database of handwritten digits," `http://yann.lecun.com/exdb/mnist/`.

LeCun, Y., Kavukcuoglu, K., and Farabet, C. (2010), "Convolutional networks and applications in vision," in *Proc. of IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 253–256.

Lee, C., Shih, J., Yu, K., and Lin, H. (2009), "Automatic music genre classification based on modulation spectral analysis of spectral and cepstral features," *IEEE Trans. Multimedia*, vol. 11, no. 4, pp. 670–682.

Lee, T. (1996), "Image representation using 2D Gabor wavelets," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 10, pp. 959–971.

Lin, J. and Qu, L. (2000), "Feature extraction based on Morlet wavelet and its application for mechanical fault diagnosis," *J. Sound Vib.*, vol. 234, no. 1, pp. 135–148.

Lowe, D. G. (2004), "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110.

Ma, J. and Plonka, G. (2010), "The curvelet transform," *IEEE Signal Process. Mag.*, vol. 27, no. 2, pp. 118–133.

Mallat, S. (2009), *A wavelet tour of signal processing: The sparse way*, Academic Press, San Diego, CA, 3rd ed.

——— (2012), "Group invariant scattering," *Comm. Pure Appl. Math.*, vol. 65, no. 10, pp. 1331–1398.

Mallat, S. and Zhong, S. (1992), "Characterization of signals from multiscale edges," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 7, pp. 710–732.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., and Hassabis, D. (2015), "Human-level control through deep reinforcement learning," *Nature*, vol. 518, pp. 529–533.

Mohamed, A., Dahl, G. E., and Hinton, G. (2011), "Acoustic modeling using deep belief networks," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 20, no. 1, pp. 14–22.

Mutch, J. and Lowe, D. (2006), "Multiclass object recognition with sparse, localized features," in *Proc. of IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11–18.

Nair, V. and Hinton, G. (2010), "Rectified linear units improve restricted Boltzmann machines," in *Proc. of International Conference on Machine Learning (ICML)*, pp. 807–814.

Naylor, A. W. and Sell, G. R. (1982), *Linear operator theory in engineering and science*, Springer, New York, NY.

Oyallon, E. and Mallat, S. (2015), "Deep roto-translation scattering for object classification," in *Proc. of IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2865–2873.

Pinto, N., Cox, D., and DiCarlo, J. (2008), "Why is real-world visual object recognition hard," *PLoS Computational Biology*, vol. 4, no. 1, pp. 151–156.

Qiao, Y. L., Song, C. Y., and Zhao, C. H. (2010), "M-band ridgelet transform based texture classification," *Pattern Recognition Letters*, vol. 31, no. 3, pp. 244–249.

Ranzato, M., Poultney, C., Chopra, S., and LeCun, Y. (2006), "Efficient learning of sparse representations with an energy-based model," in

Proc. of Int. Conf. on Neural Information Processing Systems (NIPS), pp. 1137–1144.

Ranzato, M. A., Huang, F. J., Boureau, Y., and LeCun, Y. (2007), "Unsupervised learning of invariant feature hierarchies with applications to object recognition," in *Proc. of IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–8.

Ron, A. and Shen, Z. (1995), "Frames and stable bases for shift-invariant subspaces of $L^2(\mathbb{R}^d)$," *Canad. J. Math.*, vol. 47, no. 5, pp. 1051–1094.

Rudin, W. (1983), *Real and complex analysis*, McGraw-Hill, New York, NY, 2nd ed.

——— (1991), *Functional analysis*, McGraw-Hill, New York, NY, 2nd ed.

Rumelhart, D. E., Hinton, G., and Williams, R. J. (1986), "Learning internal representations by error propagation," in J. L. McClelland and D. E. Rumelhart (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition*, pp. 318–362, MIT Press, Cambridge, MA.

Runst, T. and Sickel, W. (1996), *Sobolev spaces of fractional order, Nemytskij operators, and nonlinear partial differential equations*, vol. 3, Walter de Gruyter, Berlin.

Searcóid, M. (2007), *Metric spaces*, Springer, London.

Serre, T., Wolf, L., and Poggio, T. (2005), "Object recognition with features inspired by visual cortex," in *Proc. of IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 994–1000.

Sifre, L. (2014), *Rigid-motion scattering for texture classification*, Ph.D. thesis, Centre de Mathématiques Appliquées, École Polytechnique Paris-Saclay.

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., K., Graepel, T., and Hassabis, D. (2016), "Mastering the game of go with deep neural networks and tree search," *Nature*, vol. 529, pp. 484–489.

Simonyan, K. and Zisserman, A. (2014), "Very deep convolutional networks for large-scale image recognition," *arXiv:1409.1556*.

Tola, E., Lepetit, V., and Fua, P. (2010), "DAISY: An efficient dense descriptor applied to wide-baseline stereo," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 5, pp. 815–830.

Tzanetakis, G. and Cook, P. (2002), "Musical genre classification of audio signals," *IEEE Trans. Speech Audio Process.*, vol. 10, no. 5, pp. 293–302.

Unser, M. (1995), "Texture classification and segmentation using wavelet frames," *IEEE Trans. Image Process.*, vol. 4, no. 11, pp. 1549–1560.

Vaidyanathan, P. P. (1993), *Multirate systems and filter banks*, Prentice Hall, Englewood Cliffs, NJ.

Vandergheynst, P. (2002a), "Directional dyadic wavelet transforms: Design and algorithms," *IEEE Trans. Image Process.*, vol. 11, no. 4, pp. 363–372.

——— (2002b), "Directional dyadic wavelet transforms: Design and algorithms," *IEEE Trans. Image Process.*, vol. 11, no. 4, pp. 363–372.

Vinyals, O., Toshev, A., Bengio, S., and Erhan, D. (2015), "Show and tell: A neural image caption generator," in *Proc. of IEEE Conf. Comp. Vision Pattern Recog. (CVPR)*, pp. 3156–3164.

Viola, P. and Jones, M. J. (2004), "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154.

Waldspurger, I. (2015), *Wavelet transform modulus: Phase retrieval and scattering*, Ph.D. thesis, École Normale Supérieure, Paris.

——— (2017), "Exponential decay of scattering coefficients," *Proc. of International Conference on Sampling Theory and Applications (SampTA)*, pp. 143–146.

Wendland, H. (2004), *Scattered data approximation*, Cambridge University Press, Cambridge.

Weyl, H. (1939), "On the volume of tubes," *Amer. J. Math.*, vol. 61, pp. 461–472.

Wiatowski, T. and Bölcskei, H. (2015), "Deep convolutional neural networks based on semi-discrete frames," in *Proc. of IEEE International Symposium on Information Theory (ISIT)*, pp. 1212–1216.

——— (2018), "A mathematical theory of deep convolutional neural networks for feature extraction," *IEEE Trans. Inf. Theory*, vol. 64, no. 3, pp. 1845–1866.

Wiatowski, T., Grohs, P., and Bölcskei, H. (2017), "Topology reduction in deep convolutional feature extraction networks," in *Proc. of SPIE (Wavelets and Sparsity XVII)*, pp. 1039418:1–1039418:12.

——— (2018), "Energy propagation in deep convolutional neural networks," *IEEE Trans. Inf. Theory, to appear.*

Wiatowski, T., Tschannen, M., Stanić, A., Grohs, P., and Bölcskei, H. (2016), "Discrete deep feature extraction: A theory and new architectures," in *Proc. of International Conference on Machine Learning (ICML)*, pp. 2149–2158.

About the author

Thomas Wiatowski was born in Strzelce Opolskie, Poland, on December 20, 1987, and received the BSc in Mathematics and the MSc in Mathematics from Technical University of Munich, Germany, in 2010 and 2012, respectively. In 2012 he was a researcher at the Institute of Computational Biology at Helmholtz Zentrum in Munich, Germany. He joined ETH Zurich in 2013, where he graduated with the Dr. sc. degree in 2017. His research interests are in deep machine learning, mathematical signal processing, and applied harmonic analysis.