Wavelets

Hartmut Führ
Institut für Biomathematik und Biometrie
GSF Forschungszentrum für Umwelt und Gesundheit
email : fuehr@gsf.de

Skript zur gleichnamigen Vorlesung an der TU München
Sommersemester 2003

February 2, 2004

Introduction

This script aims at an introduction to wavelet theory. Wavelets are the result of truly interdisciplinary research, originating in the early eighties, and involving mathematicians (approximation theorists, harmonic and functional analysts), mathematical physicists as well as researchers from signal and image processing. Wavelet theory incorporated influences from all these areas of research; and nowadays wavelets can be encountered as a well-established tool in many applied fields.
This script is intended to answer the following questions:

1. What are wavelets?
2. How are wavelets constructed?
3. What are the properties of wavelets? Which additional properties are desirable?
4. Which purposes do wavelets serve?

Obviously the answers depend on the context, and complete answers cannot be given in a short script. Wavelet theory as it is developed here is essentially a subdiscipline of (Fourier) analysis, and the mathematical content consists of theorems in analysis. However the script also aims to hint why these results may be useful for nonmathematical purposes also.

There will be three (related) answers to the first question, which we will now sketch. On first reading, not everything in these short descriptions can be fully understood. The missing pieces are provided by the script. A good test for any reader will be to reread this introduction after reading the script, and to see if he/she understands it and is able to fill in the gaps.

First answer (context: image processing): Suppose we are given a *digital image*, e.g., a matrix $(g_{i,j}) \in \mathbb{R}^{2N\times 2N}$ of grey values. The **one-step fast wavelet transform associated to the Haar wavelet** is computed as follows

- Group the pixels into 2×2 subsquares $t_{i,j}$,

$$t_{i,j} = \begin{pmatrix} g_{2i,2j} & g_{2i,2j+1} \\ g_{2i+1,2j} & g_{2i+1,2j+1} \end{pmatrix}$$

- For each subsquare $t_{i,j}$ compute weighted mean values, using the weights

$$\frac{1}{4}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \quad , \quad \frac{1}{4}\begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix} \quad , \quad \frac{1}{4}\begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix} \quad , \quad \frac{1}{4}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \quad .$$

- Store these weighted mean values into $N\times N$-matrices a, d^1, d^2, d^3; hence

$$a_{1,1} = \frac{1}{4}\left(g_{1,1} + g_{1,2} + g_{2,1} + g_{2,2}\right) \quad , \quad d^1_{1,1} = \frac{1}{4}\left(g_{1,1} + g_{1,2} - g_{2,1} - g_{2,2}\right)$$

etc.

- The resulting matrices are called ”approximation matrix” (a) and detail matrices (d^1, d^2, d^3). Elements of a detail matrix are also called ”detail coefficients” or ”wavelet coefficients”.
- A standard visualisation mode for the coefficients is

$$\begin{pmatrix} a & d^1 \\ d^2 & d^3 \end{pmatrix} \quad .$$

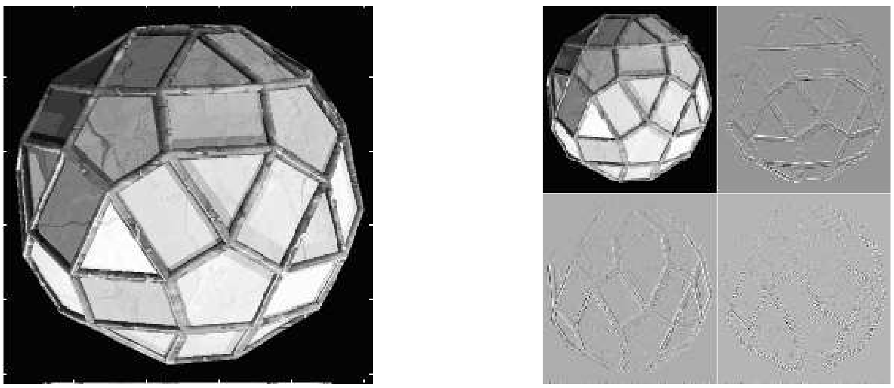

Figure 1: Original image and visualisation of approximation coefficients (left upper corner) and detail coefficients. The approximation coefficients look like a copy of the image at lower resolution, whereas the detail coefficients highlight edges of different orientations.

Iterating this procedure, where the approximation matrix is used as input for the next step, yields higher order wavelet coefficients. The usual display mode for the results is obtained by replacing the approximation coefficients by their higher order wavelet and approximation coefficients, see the next figure.

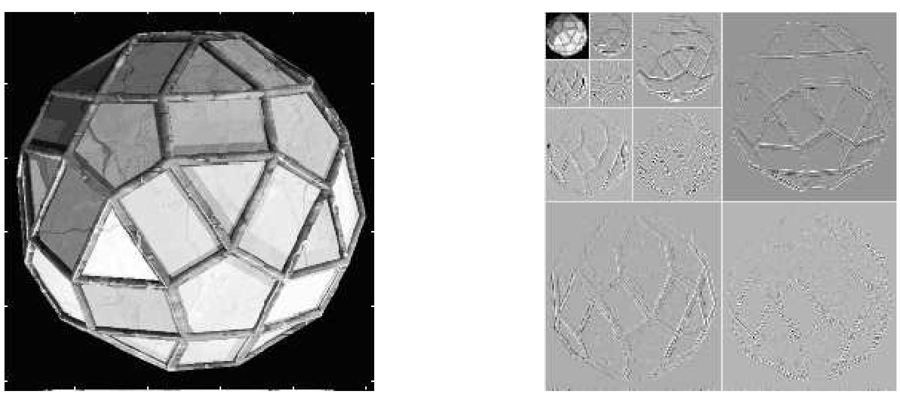

Figure 2: Original image vs. approximation coefficients (left upper corner) and detail coefficients after three iterations of the decomposition step.

The illustrations provide a fairly good intuition of the information contained in the wavelet coefficients: The approximation coefficients provide an approximation of the original image at lower resolution. Depending on the orientation (that is, the index i of d^i), the wavelet coefficients d^i are large at positions close to abrupt changes in the image along the vertical or horizontal direction.

Let us collect the main properties of the wavelet coefficients:

- The computation of wavelet coefficients is simple and fast.
- Likewise, there is a simple and fast algorithm for the reconstruction of the image from wavelet coefficients.
- The coefficients corresponding to image regions with almost constant grey levels are small (this turns out to be particularly useful for compression purposes).

Second answer (analysis): The continuous wavelet transform on $\mathbb{R}$ is a linear mapping assigning functions on $\mathbb{R}$ certain functions on $\mathbb{R} \times \mathbb{R}^*$, by the following procedure: Suppose we are given a function $g : \mathbb{R} \to \mathbb{R}$, which is smooth with the exception of a few isolated points.

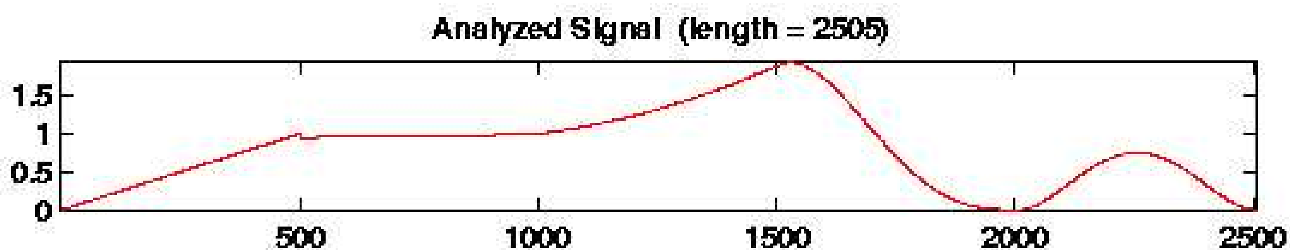

Figure 3: Input signal f. f or one of its derivatives has jumps at positions 500,100,1500, 2000.

We define

$$\psi(x) = (1 - x^2)e^{-x^2/2} \quad ,$$

as well as for $b \in \mathbb{R}$ and $a > 0$,

$$\psi_{b,a} = |a|^{-1/2}\psi\left(\frac{x-b}{a}\right)$$

a version of ψ, **scaled** by a and **shifted** by b. A popular choice is plotted below.

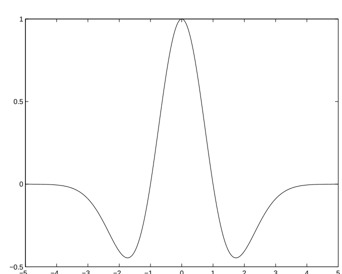

Figure 4: The mexican hat wavelet.

The **continuous wavelet transform** of g is defined as the map $W_\psi g : \mathbb{R} \times \mathbb{R}^+ \to \mathbb{R}$ given by

$$\begin{aligned} W_\psi g(b,a) &= \langle g, \psi_{b,a} \rangle \\ &= \int_{\mathbb{R}} g(x)\psi\left(\frac{x-b}{a}\right) dx \end{aligned}$$

A plot of the wavelet coefficients shows that large coefficients concentrate in cones pointing at the "jumps" of g.

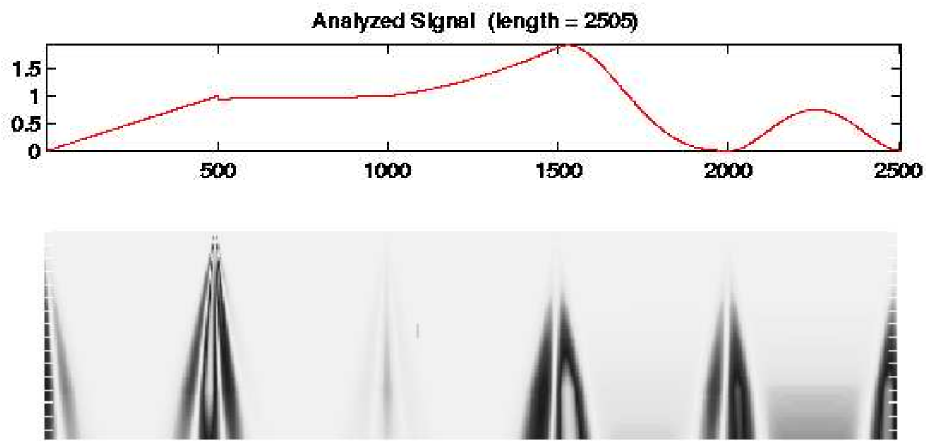

Figure 5: Wavelet coefficients $W_\psi(b,a)$. The horizontal axis corresponds to the position parameter b, the vertical axis to scale (inverted, i.e., lower points on the axis correspond to larger scales). Dark colors correspond to large absolute values; these tend to accumulate around the jump points.

The common features of the two different notions of a wavelet turn out to be:

- "Wavelets" are elementary building blocks indexed by a scale and a position parameter.
- A "wavelet transform" consists in a decomposition of a "signal" (image, function) into these building blocks, by computing certain expansion coefficients, called "wavelet coefficients". The expansion coefficients are scalar products of the signal with the building blocks. There is also an inverse map allowing to "synthesise" the signal from the coefficients.
- Large wavelet coefficients occur mostly in the vicinity of singularities.

Besides these somewhat hazy analogies between the two answers, there also exists a mathematical connection, which will be the

Third answer (orthonormal wavelets): An **orthonormal wavelet** is a function $\psi \in \mathrm{L}^2(\mathbb{R})$ with the property that

$$\psi_{j,k}(x) = 2^{j/2}\psi\left(2^j t - k\right)$$

defines an orthonormal basis $(\psi_{j,k})_{j,k\in\mathbb{Z}}$ of $\mathrm{L}^2(\mathbb{R})$.

Note that it is not apparent at all why the third answer connects the two previous ones; except maybe that wavelet ONB's are discretely indexed.

The chief aims of this script are:

- Give a mathematical treatment of the second and third answer. Study the existence and properties of continuous and discrete wavelet systems in $\mathrm{L}^2(\mathbb{R})$, as well as construction methods.

- Derive the fast wavelet transform from the construction of wavelet orthonormal bases via **multiresolution analyses**. Sketch wavelet based algorithms for certain signal processing problems, and explain the heuristics behind these algorithms.

Outline

The first chapter recalls basics from analysis. Wavelet analysis in general is about decomposing functions into basic building blocks. These building blocks are indexed by a scale and a translation parameter. The first chapter clarifies the different ways of decomposing a function; mathematically this is expressed by certain isometric operators between Hilbert spaces. A particularly intuitive type of decomposition is provided by orthonormal bases (ONB's), but we will also need "continuous" versions (such as the Fourier transforms), which are best understood by analogy to ONB's. The basics concerning Hilbert spaces are presented in the first two section. Then we give an overview of integration and L^p-spaces. Here the most important aspects concern pointwise statements about L^p-functions (there are a few traps one has to avoid), and convergence in L^p. We also define the class of "test signals" which provide a sort of benchmark to compare different transforms, namely piecewise smooth functions.

The last three sections are concerned with the Fourier transform and its properties. The role of the Fourier transform in this script is twofold: On the one hand, it is impossible to prove anything about wavelets without it. On the other hand the Fourier transform serves as a yardstick for wavelets: We will often compare wavelets (favorably) to the Fourier transform. Wavelets were designed to remedy certain shortcomings of the Fourier transform, and one of them can be understood by use of piecewise smooth functions: The global smoothness of a functions is closely related to the decay behaviour of its Fourier transform. However, if the function is very smooth *except at one point*, this close relation vanishes. This can be attributed to the "lack of localisation" of the Fourier transform: Changing the function locally has a global effect on its Fourier transform. These observations are explained in detail in Sections 1.5 and 1.6.

Chapter 2 deals with the continuous wavelet transform (CWT, the second of the above answers). The CWT is in a sense the simplest transform to construct. One should keep in mind that generally a wavelet is a tool that can be chosen according to the need of the user. For the CWT, the arsenal of wavelets to choose from is particularly large; they only have to fulfill a (rather mild) admissibility condition. We then discuss interpretation issues. The scale parameter can be read as inverse frequency, which allows to read the CWT as something like a "localised Fourier transform" or "filterbank". This interpretation will provide a good intuition for the analysis of piecewise smooth functions. Moreover, it gives a first hint that a good wavelet is one that is "well-localised in time and frequency", i.e., with fast decay of the wavelet and its Fourier transform. We then deal with piecewise smooth functions and their characterisation using the decay behavior of the CWT (Sections 2.2 through 2.4). As a natural byproduct we obtain a more precise formulation of the rather hazy "good localisation" requirement, namely in terms of decay behaviour, vanishing moments and smoothness of the wavelet.

We also want to draw the reader's attention to two wavelets which keep popping up throughout the script, namely the Haar and the Shannon wavelet. They are somewhat extremal and academic cases which serve mainly as illustration of the various phenomena.

However, at least the Haar wavelet is still often encountered in applications, and very often the heuristic arguments used for certain wavelet based algorithms can be best understood by thinking of the Shannon wavelet. In particular, the view of wavelet transform as filterbank separating different frequency components is valid in a sharp sense only for the Shannon wavelet.

In Chapter 3 we consider wavelet ONB's of $L^2(\mathbb{R})$, and their construction via multiresolution analyses. In a sense, the course of argument consists in "boiling down" the construction problem in several steps: First, the construction of an MRA (which is a sequence of closed subspaces of $L^2(\mathbb{R})$) is seen to correspond to the correct choice of a single function $\varphi L^2(\mathbb{R})$, the *scaling function* (Section 3.1). Then the scaling function is seen to be characterised by a certain discrete series of coefficients (called "scaling coefficients") (Section 3.3). Hence a clever choice of these coefficients is enough to construct the MRA (and the wavelet ONB is a simple byproduct of this, see Section 3.2). That such a clever choice of coefficients can be made in such a way as to yield compactly supported wavelets with arbitrarily many vanishing moments and arbitrary smoothness, is a famous result due to Daubechies. We give a sketch of the proof in Section 3.3. In Section 3.4 we outline an alternative approach to wavelet construction, which results in spline wavelets.

Chapter 4 deals with purely discrete data. As a byproduct of the multiresolution construction we obtain a fast and simple algorithm which computes wavelet coefficients of a function $L^2(\mathbb{R})$ its sequence of scaling coefficients. The importance of the algorithm for the success of wavelets in applications cannot be overstated. The algorithm implements an operator on the (discrete time) signal space $\ell^2(\mathbb{Z})$. Due to its origin, the operator has an interpretation in connection with an MRA in $L^2(\mathbb{R})$. However, it is also possible to understand it as a strictly discrete-time filterbank (watch out for the Shannon wavelet), or as computing the expansion coefficients with respect to a certain ONB of $\ell^2(\mathbb{Z})$.

In Section 4.3 we address the extension of the algorithm to 2D signals, i.e. discrete images. In the final section we comment on two standard applications of discrete wavelets, namely denoising and compression.

Sources

There are numerous books on wavelets, but unfortunately none which contains (in an easily accessible manner) the material presented here. Moreover, as anyone trying to compare two wavelet books knows, every author uses his own indexing of MRA's and has his own ideas where to put the 2π in the Fourier transform etc. Hence, if nothing else, this script provides a self-contained account of the several wavelet transforms and their connections.

Chapter 1 covers results contained in any introductory book on real and/or functional analysis. I only give proofs in Chapter 1 when I regard them as useful for the further understanding of the script.

Chapter 2 is uses ideas and results from [4, 3, 14], though I'm not aware of a source containing rigourous proofs of Theorems 2.2.13 or 2.3.3. There exist proofs for related (usually rather simpler) statements, e.g., in [5, 1].

Chapter 3 is mostly taken from [16]. Chapter 4 uses parts from [2, 6].

For further reading, we recommend [6] for a rather complete account of signal processing applications, [9, 10] for the use of wavelet transforms in functional and Fourier analysis, and

[11] for a good overview of wavelets, time-frequency analysis and the like.

All illustrations in this script were produced using the Matlab wavelets toolbox.

Contents

Chapter 1

Basics of (Functional) Analysis

1.1 Hilbert spaces and ONB's

As explained in the introduction, the decomposition of functions into building blocks is a major issue in this script. The suitable environment for this type of problem is provided by **Hilbert spaces**. In order to motivate Hilbert spaces, let us recall a few facts from linear algebra:

A standard application of Zorn's lemma provides the existence of an (algebraic) basis $(x_i)_{i\in I}$, for every vector space X. Hence each element $x \in X$ can be written as a unique linear combination

$$x = \sum_{i\in I} \alpha_i x_i$$

with only finitely many nonlinear coefficients. However, we are interested in vector spaces of functions, and these spaces are usually infinite-dimensional; the algebraic dimension of decent functions spaces is in fact uncountable. This causes at least two unpleasant problems:

- Bases in the sense of linear algebra exist, but they are impossible to compute explicitly (Zorn's lemma is not constructive).
- Even if we could somehow compute a basis, the computation of the coefficients would still require solving an infinite-dimensional system of linear equations (forget it).

One way around both problems (in fact, the simplest one that works) is provided by Hilbert spaces and orthonormal bases (ONB's). Most interesting Hilbert spaces contain explicitly computable countable orthonormal bases $(x_i)_{i\in I}$ (the wavelet ONB's in $\mathrm{L}^2(\mathbb{R})$, constructed in chapter 3, are one class of examples). And by orthonormality, computing the coefficients amounts to taking scalar products. In short, the decomposition provided by orthonormal bases is given

$$x = \sum_{i\in I} \langle x, x_i \rangle x_i \quad ,$$

where we used the scalar product of the Hilbert space.

Definition 1.1.1 Let $\mathcal{H}$ be a $\mathbb{C}$-vector space. A mapping

$$\langle \cdot, \cdot \rangle : \mathcal{H} \times \mathcal{H} \to \mathbb{C}$$

is a **scalar product**, when the following statements hold:

1. For all $x \in \mathcal{H}$ the mapping $y \mapsto \langle y, x \rangle$ is linear.
2. For all $x, y \in \mathcal{H}$: $\langle x, y \rangle = \overline{\langle y, x \rangle}$.
3. For all $x \in \mathcal{H}$, $\langle x, x \rangle \geq 0$, with equality only for $x = 0$.

We define the **norm** on $\mathcal{H}$ as $\|x\| := \sqrt{\langle x, x \rangle}$. A space $\mathcal{H}$ endowed with scalar product is called pre-Hilbert space. □

1.1.2 Lemma and Definition. Let $\mathcal{H}$ be a pre-Hilbert space.

(a) **Cauchy-Schwarz inequality**

$$\forall x, y \in \mathcal{H} \; : \; |\langle x, y \rangle| \leq \|x\| \|y\| \quad .$$

(b) **Triangle inequality**

$$\forall x, y \in \mathcal{H} \; : \; \|x + y\| \leq \|x\| + \|y\|$$

(c) The mapping $d(x, y) := \|x - y\|$ is a metric on $\mathcal{H}$.

If $\mathcal{H}$ is complete with respect to the metric d, it is called **Hilbert space**. □

Definition 1.1.3 (Unconditional convergence)
Let $\mathcal{H}$ be a Hilbert space, $(x_i)_{i \in I} \subset \mathcal{H}$ a family of vectors, and $x \in \mathcal{H}$. We write

$$x = \sum_{i \in I} x_i$$

if for all $\epsilon > 0$ there exists a finite set $J_\epsilon \subset I$ such that for all finite $J' \supset J_\epsilon$,

$$\left\| x - \sum_{i \in J'} x_i \right\| < \epsilon \quad .$$

□

Example 1.1.4 (a) Let I be a set. Define

$$\ell^2(I) = \{(a_i)_{i \in I} \in \mathbb{C}^I : \sum_{i \in I} |a_i|^2 < \infty\} \quad .$$

Then $\ell^2(I)$ is a Hilbert space, with scalar product

$$\langle (a_i)_{i \in I}, (b_i)_{i \in I} \rangle = \sum_{i \in I} a_i \overline{b_i} \quad .$$

(See Section 1.3 below for the more general setting of L^2-spaces.)
(b) Let $\mathcal{H}$ be a Hilbert space and $\mathcal{H}' \subset \mathcal{H}$ a closed subspace (closed with respect to the metric, that is). Then $\mathcal{H}'$ is a Hilbert space.
(c) Let $\mathcal{H}_1$, $\mathcal{H}_2$ be Hilbert spaces. Then the **orthogonal sum** $\mathcal{H}_1 \oplus \mathcal{H}_2$ is the direct product of $\mathcal{H}_1$ and $\mathcal{H}_2$, endowed with the scalar product

$$\langle (x_1, x_2), (y_1, y_2) \rangle = \langle x_1, y_1 \rangle_1 + \langle x_2, y_2 \rangle_2 \quad .$$

Here $\langle \cdot, \cdot \rangle_i$ denotes the scalar product on the Hilbert space $\mathcal{H}_i$. It is easy to check that the orthogonal sum is complete, i.e., $\mathcal{H}_1 \oplus \mathcal{H}_2$ is a Hilbert space. □

Definition 1.1.5 Let $\mathcal{H}$ be a Hilbert space.

(a) Two vectors x, y are called orthogonal ($x \perp y$) if $\langle x, y \rangle = 0$.

(b) A family $(x_i)_{i\in I} \subset \mathcal{H}$ is an **orthonormal system** (ONS) if

$$\forall i, j \in I \ : \ \langle x_i, x_j \rangle = \delta_{i,j}$$

(c) A family $(x_i)_{i\in I} \subset \mathcal{H}$ is **total** if for all $x \in \mathcal{H}$ the following implication holds

$$(\forall i \in I \ : \ \langle x, x_i \rangle = 0) \Rightarrow x = 0 \ .$$

(d) A total ONS is called **orthonormal basis** (ONB).

□

Proposition 1.1.6 *Let $\mathcal{H}$ be a Hilbert space.*

(a) (Pythagoras) If $x \perp y$ then $\|x + y\|^2 = \|x\|^2 + \|y\|^2$.

(b) (Parseval equality) Let $(x_i)_{i\in I}$ be an ONS, $(a_i)_{i\in I} \in \ell^2(I)$. Then $\sum_{i\in I} a_i x_i$ converges, with $\| \sum_{i\in I} a_i x_i \|^2 = \sum_{i\in I} |a_i|^2$.

(c) Let $(x_i)_{i\in I}$ be an ONS, $x \in \mathcal{H}$. Then $(\langle x, x_i \rangle)_{i\in I} \in \ell^2(I)$, with $\sum_{i\in I} |\langle x, x_i \rangle|^2 \le \|x\|^2$.

(d) Let $(x_i)_{i\in I}$ be an ONS. Suppose that $x = \sum_{i\in I} a_i x_i$, with $(a_i)_{i\in I} \in \ell^2(I)$. Then $a_i = \langle x, x_i \rangle$.

Theorem 1.1.7 *Let $(x_i)_{i\in I}$ be an ONS in the Hilbert space $\mathcal{H}$. Then the following are equivalent:*

(a) $(x_i)_{i\in I}$ ONB.

(b) For all $x \in \mathcal{H}$: $x = \sum_{i\in I} \langle x, x_i \rangle x_i$.

1.1.8 Exercise
(a) Show that every Hilbert space contains an ONB. (Use Zorn's lemma.)
(b) Let $x = \sum_{i\in I} x_i$ and $y \in \mathcal{H}$. Prove that $\langle x, y \rangle = \sum_{i\in I} \langle x_i, y \rangle$, where the right hand side converges unconditionally. □

Remark 1.1.9 If $(x_i)_{i\in I}$ is an ONS, then

$$(x_i)_{i\in I} \text{ ONB } \Leftrightarrow \forall x \in \mathcal{H} \ : \|x\|^2 = \sum_{i\in I} |\langle x, x_i \rangle|^2$$

□

1.1.10 Exercise: In $\mathbb{C}$ unconditional convergence is the same as absolute convergence. Let $(x_i)_{i\in I} \subset \mathbb{C}$. Then $\sum_{i\in I} x_i$ converges in the sense of 1.1.3 iff $\sum_{i\in I} |x_i| < \infty$-
□

1.2 Linear operators and subspaces

Definition 1.2.1 Let $\mathcal{H}, \mathcal{H}'$ be Hilbert spaces, $T : \mathcal{H} \to \mathcal{H}'$ a linear operator. T is called **bounded**, if $\|Tx\| \le c_T \|x\|$ holds, for all $x \in \mathcal{H}$ and a fixed constant $c_T > 0$. A bounded linear operator $\mathcal{H} \to \mathbb{C}$ is called **linear functional**. □

Remark 1.2.2 An linear operator is bounded iff it is continuous with respect to the metrics on $\mathcal{H}$ and $\mathcal{H}'$. (The "only if"-part is straightforward, the "if"-part is treated in every functional analysis course.) □

Theorem 1.2.3 *(Fischer/Riesz)*
Let $\mathcal{H}$ be a Hilbert space. Then, for all $x \in \mathcal{H}$ the mapping $y \mapsto \langle y, x \rangle$ is a linear functional. Conversely, if $\varphi : \mathcal{H} \to \mathbb{C}$ is a linear functional, there exists $x \in \mathcal{H}$ such that for all $y \in \mathcal{H}$: $\varphi(y) = \langle y, x \rangle$.

Proof. The fact that every x gives rise to a linear functional is due to the Cauchy-Schwarz inequality. The converse is again covered in functional analysis. □

1.2.4 Theorem and Definition Let $T : \mathcal{H} \to \mathcal{H}'$ be a bounded linear operator. Then there exists a unique bounded linear operator $T^* : \mathcal{H}' \to \mathcal{H}$ such that for all $x \in \mathcal{H}$ and all $y \in \mathcal{H}'$

$$\langle Tx, y \rangle = \langle x, T^* y \rangle \quad .$$

T^* is called the **adjoint operator** of T. □

Definition 1.2.5 Let $\mathcal{H}$ and $\mathcal{H}'$ be Hilbert spaces and $T : \mathcal{H} \to \mathcal{H}'$.

(a) T is an **(orthogonal) projection** if $T = T^* = T^2$.

(b) T is **unitary** if $TT^* = \mathrm{Id}_{\mathcal{H}}$ and $T^*T = \mathrm{Id}_{\mathcal{H}'}$.

(c) T is **isometric** if $\|Tx\| = \|x\|$, for all $x \in \mathcal{H}$.

□

Proposition 1.2.6 *Let $\mathcal{H}$ be a Hilbert space, and $\mathcal{K} \subset \mathcal{H}$ a subset.*

(a) $\mathcal{K}^\perp = \{z \in \mathcal{H} : (\forall x \in \mathcal{H} : \langle z, x \rangle = 0)\}$ is a closed subspace.

(b) Assume that $\mathcal{K}$ is a closed subspace. Then for all $x \in \mathcal{H}$ there exists a unique pair $(y, y^\perp) \in \mathcal{K} \times \mathcal{K}^\perp$ such that $x = y + y^\perp$. The mappings $x \mapsto y$ and $x \mapsto y^\perp$ are projections ("onto $\mathcal{K}$" resp. "onto $\mathcal{K}^\perp$).

(c) Conversely: If $T : \mathcal{H} \to \mathcal{H}$ is a projection, then $\mathcal{K} = T(\mathcal{H})$ is a closed subspace. The projection onto $\mathcal{K}^\perp$ is given by $\mathcal{I}\lceil_{\mathcal{H}} - T$.

1.2.7 Exercises. Let $T : \mathcal{H} \to \mathcal{H}'$ be a bounded linear operator between two Hilbert spaces.

(a) T is isometric iff $\langle Tx, Ty \rangle = \langle x, y \rangle$ for all $x, y \in \mathcal{H}$.
Hint: Prove and apply the **polarisation equation**

$$\langle x, y \rangle = \frac{1}{4} \left(\|x + y\|^2 - \|x - y\|^2 + i(\|x - iy\|^2 + \|x + iy\|^2) \right) \quad .$$

(b) T is isometric iff TT^* is a projection and $T^*T = \mathrm{Id}_{\mathcal{H}}$.

$\square$

Corollary 1.2.8 *For $(x_i)_{i\in I} \subset \mathcal{H}$ the following are equivalent.*

(a) $(x_i)_{i\in I}$ *is total.*

(b) $\{x_i : i \in I\}^\perp = \{0\}$.

(c) $\overline{\mathrm{span}\{x_i : i \in I\}} = \mathcal{H}$.

Lemma 1.2.9 *Let $T : \mathcal{H} \to \mathcal{H}'$ be a bounded operator between Hilbert spaces.*

(a) If T is isometric, then it is injective.

(b) T is unitary iff T is isometric and onto.

Corollary 1.2.10 *Let $(x_i)_{i\in I} \subset \mathcal{H}$ be an ONS. Then the* **coefficient operator** $T : \mathcal{H} \to \ell^2(I)$, $Tx = (\langle x, x_i \rangle)_{i\in I}$ *is well-defined, bounded and onto. The following are equivalent:*

(a) T is isometric.

(b) T is unitary.

(c) $(x_i)_{i\in I}$ is an ONB.

Remark 1.2.11 One important aspect of ONBs in finite dimensional spaces is that they induce a coordinate system. Recall from linear algebra that a clever choice of coordinate system can simplify certain problems considerably. For instance, the spectral theorem implies that a selfadjoint operator is diagonal in a suitable basis. Hence in this coordinate system properties of the operator such as invertibility, positive definiteness etc. can be immediately read off the matrix of the operator with respect to the eigenbasis. (This observation generalises to Hilbert spaces.)

This idea of choosing a clever basis can be seen as one motivation of wavelets: As the discussion in Section 1.6 shows, it is not easy to decide whether a function is piecewise smooth by looking at its "Fourier coordinates", whereas Chapter 2 shows that "wavelet coordinates" can be employed for this purpose. $\square$

1.3 Function spaces

In this section we recall the basic notions of integration theory, as needed for the following. Notions and results not mentioned here can be found e.g. in [13].

Definition 1.3.1 Let $(X, \mathcal{B}, \mu)$ be a measure space, i.e.,

- a set X
- a σ-algebra $\mathcal{B}$
- a **measure** μ, i.e. a σ-additive map $\mu : \mathcal{B} \to \mathbb{R}_0^+ \cup \{\infty\}$.

We define $M(X)$ as the vector space of measurable functions $X \to \mathbb{C}$, and $N(X)$ as the vector space of almost everywhere vanishing functions, i.e.,

$$N(X) = \{f \in M(X) : X \to \mathbb{C} : \exists M \in \mathcal{B} \text{ such that } \mu(B) = 0, f(x) = 0 \forall x \notin B \ \ \} \ \ .$$

For $1 \le p < \infty$, the space $\mathrm{L}^p(X,\mu) = \mathrm{L}^p(X)$ is defined as the quotient space

$$\mathrm{L}^p(X) = \{f \in M(X) : \int_X |f(x)|^p d\mu(x) < \infty\}/N(X)$$

endowed with the L^p-norm

$$\|f\|_p = \left(\int_X |f(x)|^p d\mu(x) \right)^{1/p}$$

For $p = \infty$ we let

$$\mathrm{L}^\infty(X) = \{f \in M(X) : \exists g \in M(X), g \text{ bounded}, f - g \in N(X)\}/N(X)$$

with

$$\|f\|_\infty = \inf\{\alpha : |f(x)| < \alpha \text{ almost everywhere } \}$$

□

The main reason for writing down the definition of L^p-spaces is to draw attention to the fact that L^p-functions are in fact equivalence classes. It is customary to use a somewhat ambiguous language which refers both to equivalence classes and their representatives as L^p-functions.

Example 1.3.2 (a) Let $X \subset \mathbb{R}^n$ be a Borel set, and $\mathcal{B}$ the σ-algebra of Borel subsets of X. Let λ be Lebesgue measure on $\mathbb{R}^n$, $\varphi : X \to \mathbb{R}_0^+$ a measurable function, and consider the measure

$$\mu(A) = \int_X \varphi(x)dx \ \ .$$

The corresponding L^p-space is denoted by $\mathrm{L}^p(X, \phi(x)dx)$.
(b) Let $X = I$, an arbitrary set, and $\mathcal{B}$ the power set of X, and consider the counting measure μ on X. Then $\ell^2(I) = \mathrm{L}^2(X,\mu)$. □

Theorem 1.3.3 *Let $(X, \mathcal{B}, \mu)$ be a measure space, $1 \le p \le \infty$. Then $\mathrm{L}^p(X)$ is a* **normed space**, *i.e., $\| \cdot \|_p$ fulfills*

- $\|g\|_p = 0 \Rightarrow g = 0$
- $\|\alpha g\|_p = |\alpha| \ \|g\|_p$
- $\|f + g\|_p \le \|f\|_p + \|g\|_p$

Moreover, $\mathrm{L}^p(X)$ is complete with respect to the metric $d(f,g) = \|f - g\|_p$; this makes $\mathrm{L}^p(X)$ a **Banach space**. *$\mathrm{L}^2(X)$ is a Hilbert space, with scalar product*

$$\langle f, g \rangle = \int_X f(x)\overline{g(x)}d\mu(x) \ \ .$$

In the script we will discuss pointwise properties of L^p-functions, such as differentiability. This seems to contradict the fact that elements of L^p are equivalence classes of measurable functions modulo almost everywhere vanishing functions. It is not a priori obvious that in this setting smoothness is a useful notion: One can take an arbitrarily smooth function, add to this function the characteristic function of the rationals, and obtain something that is nowhere continuous, but in the same equivalence class mod $N(X)$.

Definition 1.3.4 A **piecewise C^n-function** is a pair $(f, S(f))$, where $f : \mathbb{R} \to \mathbb{C}$, $S(f) \subset \mathbb{R}$ is discrete (i.e. without accumulation point), such that f is n times continuously differentiable in $\mathbb{R} \setminus S(f)$. $S(f)$ is called the set of singularities of f. We call $f \in \mathrm{L}^p(\mathbb{R})$ piecewise C^n if there exists a representative of f which is piecewise C^n. □

The following proposition shows that speaking of piecewise C^n L^p-functions does make sense. In fact, there exists a "canonical" representative with a minimal set of singularities.

1.3.5 Proposition and Definition Let $f \in \mathrm{L}^p(\mathbb{R})$ be piecewise C^n. Then f has a piecewise C^n representative $(g, S(g))$ with the property that for every other piecewise C^n representative $(h, S(h))$ we have $S(g) \subset S(h)$. This property fixes g uniquely on g. g is called "canonical representative". □

1.3.6 Remark. (Pointwise convergence vs. norm convergence) Convergence in $\|\cdot\|_p$ does **not** imply pointwise convergence, not even almost everywhere. But at least we have

$$\|f_n - f\| \to 0 \Rightarrow \exists \text{ subsequence } (f_{n_k})_{k\in\mathbb{N}} \text{ such that } f_{n_k}(x) \to f(x) \text{ almost everywhere} \quad .$$

The converse implication is even more problematic: $f_n \to f$ pointwise usually does not imply $f_n \to f$ in the norm. However, there are additional criteria to ensure norm convergence:

(a) (Monotone Convergence) If $0 \le f_1 \le f_2 \le \ldots$, then

$$\int_X f(x)dx = \lim_{n\to\infty} f_n(x)dx$$

and

$$\|f_n - f\|_1 \to 0 \Leftrightarrow f \in \mathrm{L}^1(X) \quad .$$

(b) (Dominated Convergence) If $|f_n(x)| \le g(x)$ for all $x \in X$ and a fixed $g \in \mathrm{L}^1(X)$, then $f \in \mathrm{L}^1(X)$,

$$\int_X f(x)dx = \lim_{n\to\infty} f_n(x)dx$$

and $\|f - f_n\|_1 \to 0$.

(c) If $f_n \to f$ uniformly, and if f_n and f vanish outside a fixed set of finite measure, then $f_n \in \mathrm{L}^p(X)$ implies $f \in \mathrm{L}^p(X)$ and $f_n \to f$ in $\|\cdot\|_p$.

□

1.3.7 Dense subspaces of L^p Let $1 \le p < \infty$. The following subspaces are dense in $\mathrm{L}^p(\mathbb{R}^n)$:

(a) The space of **step functions**,

$$T(\mathbb{R}^n) = \{\sum_{i=1}^{m} \alpha_i \chi_{C_i} \ : \ m \in \mathbb{N}, \alpha_i \in \mathbb{C}, C_i = [a_{1,i}, b_{1,i}] \times \ldots [a_{n,i}, b_{n,i}]\}$$

(b) $C_c^\infty(\mathbb{R}^n) = \{f : \mathbb{R}^n \to \mathbb{C} : \text{supp}(f) \text{ compact, and } \forall m \in \mathbb{N} : f \in C^m(\mathbb{R}^n)\}$ is dense. Moreover, if $I \subset \mathbb{R}^n$ is open, then $\{f \in C_c^\infty(\mathbb{R}^n) : \text{supp} f \subset I\}$ is dense in $\mathrm{L}^p(I)$.

□

1.3.8 Proposition. (Inclusion between L^p-spaces.) Let $(X, \mathcal{B}, \mu)$ be a measure space.

(a) If $\mu(X) < \infty$ and $1 \leq q \leq p \leq \infty$, then $\mathrm{L}^p(X) \subset \mathrm{L}^q(X)$.

(b) For all $1 \leq q \leq p \leq \infty$ we have $\mathrm{L}^p(X) \supset \mathrm{L}^q(X) \cap \mathrm{L}^\infty(X)$. In particular, $\ell^p(X) \supset \ell^q(X)$.

□

1.3.9 Exercise. (The Haar basis)

(a) Define $\psi = \chi_{[0,1/2[} - \chi_{[1/2,1]}$, and, for $j,k \in \mathbb{Z}$

$$\psi_{j,k}(t) = 2^{j/2}\psi(2^j(t-k)) \quad .$$

Moreover let $\varphi = \chi_{[0,1]}$. Prove that $\{\varphi\} \cup \{\psi_{j,k} : j \geq 0, 0 \leq k \leq 2^j - 1\}$ is an ONB of $\mathrm{L}^2([0,1])$.
(Hint: Orthonormality is straightforward. For totality, first prove for all $j_0 \geq 0$ that the span of $\{\varphi\} \cup \{\psi_{j,k} : j_0 \geq j \geq 0, 0 \leq k \leq 2^j - 1\}$ is given by the functions $f : [0,1] \to \mathbb{C}$ which are constant on the "dyadic intervals" $[2^{-j_0-1}k, 2^{-j_0-1}(k+1)]$.)

(b) Show that $\{\psi_{j,k} : j,k \in \mathbb{Z}\}$ is an ONB of $\mathrm{L}^2(\mathbb{R})$.
(Hint: Again, totality is the nontrivial part. For this purpose use part (a) to show that for all $j_0 \in \mathbb{Z}$ the family $(\psi_{j,k})_{j \geq j_0, 0 \leq k \leq 2^{j-j_0}-1}$ is an ONB of

$$\mathcal{H}_{j_0} = \{f \in \mathrm{L}^2([0, 2^{-j_0}]) : \int_0^{2^{-j_0}} f(x)dx = 0\} \quad .$$

Then prove that $\bigcup_{j_0=1}^{-\infty} \mathcal{H}_{j_0}$ is dense in $\mathrm{L}^2([0,\infty[)$. A simple reflection argument yields the same for $\mathrm{L}^2(]-\infty,0])$.)

□

1.4 Fourier Series

Theorem 1.4.1 *For $T > 0$ consider the Hilbert space $\mathcal{H} = \mathrm{L}^2([-T/2, T/2])$ and the sequence $(e_n)_{n\in\mathbb{Z}} \subset \mathcal{H}$ defined by*

$$e_n(t) = T^{-1/2} e^{2\pi i n t/T} \quad .$$

Then $(e_n)_{n\in\mathbb{Z}}$ is an ONB of $\mathcal{H}$, called the **Fourier basis**.

As a consequence of the theorem we obtain the **Fourier transform** $\mathcal{F} : \mathrm{L}^2([-T/2, T/2]) \to \ell^2(\mathbb{Z})$, which is the coefficient operator with respect to the Fourier basis. By 1.2.10 the Fourier transform is a unitary operator. Every $f \in \mathrm{L}^2([-T/2, T/2])$ can be expanded in the **Fourier series**

$$f = \sum_{n \in \mathbb{Z}} \langle f, e_n \rangle e_n \quad .$$

We use the notation $\widehat{f}(n) = \langle f, e_n \rangle$. The Fourier transform may be extended to the larger space $L^1([-T/2, T/2])$, retaining at least some properties:

Theorem 1.4.2 *For $f \in \mathrm{L}^1([-T/2, T/2])$ and $n \in \mathbb{Z}$, the right hand side of*

$$\widehat{f}(n) = T^{-1/2} \int_{-T/2}^{T/2} f(t) e^{-2\pi i n t/T} dt$$

is absolutely convergent. The map $f \mapsto \widehat{f} = (\widehat{f}(n))_{n \in \mathbb{Z}}$ is a linear operator $\mathcal{F} : \mathrm{L}^1([-T/2, T/2]) \to c_0(\mathbb{Z})$, called the **($\mathrm{L}^1$-) Fourier transform**. *($c_0(\mathbb{Z})$ denotes the space of sequences converging to zero). The operator $\mathcal{F}$ is injective.*

Remark 1.4.3 The set $[-T/2, T/2]$ may be replaced by any measurable set $A \subset \mathbb{R}$ with the properties

- $A \cap nT + A$ has measure zero, for all $n \in \mathbb{Z}$. (Here $nT + A = \{nT + a : a \in A\}$.)
- $\mathbb{R} \setminus \left(\bigcup_{n \in \mathbb{Z}} nT + A \right)$ has measure zero.

□

1.5 Fourier integrals

An informal derivation of Fourier integrals from Fourier series can be obtained as follows: Let $f \in \mathrm{L}^2([-T/2, T/2])$. Define for $\omega \in \mathbb{R}$ the function

$$e_\omega(x) = e^{2\pi i \omega x} \quad (x \in \mathbb{R}) \quad .$$

For $S > T$ we can write (at least in the L^2-sense)

$$\begin{aligned} f(x) &= \sum_{n \in \mathbb{Z}} \int_{-S/2}^{S/2} f(t) S^{-1/2} e^{-2\pi i t n/S} dt S^{-1/2} e^{2\pi i x n/S} \\ &= \sum_{n \in \mathbb{Z}} \langle f, e_{n/S} \rangle \frac{e_{n/S}(x)}{S} \\ &\to \int_{\mathbb{R}} \langle f, e_\omega \rangle e_\omega(x) d\omega \quad (S \to \infty), \end{aligned}$$

where the last line is a strictly formal observation that the Riemann sums should converge to the integral. This is obviously highly nonrigourous; in particular, the e_ω are not in $\mathrm{L}^2(\mathbb{R})$. A rigourous definition requires a detour via $\mathrm{L}^1(\mathbb{R})$:

Theorem 1.5.1 *Let $f \in \mathrm{L}^1(\mathbb{R})$ be given, and define*

$$\widehat{f}(\omega) = \int_{\mathbb{R}} f(t) e^{-2\pi i \omega t} dt \quad (\omega \in \mathbb{R}).$$

The **(L^1-) Fourier transform** *is the linear operator $\mathcal{F} : f \mapsto \widehat{f}$.*

(a) $\widehat{f}$ is continuous and bounded, with $\|\widehat{f}\|_\infty \leq \|f\|_1$.

(b) **Riemann-Lebesgue:** $\lim_{|\omega| \to \infty} \widehat{f}(\omega) = 0$.

(c) **Fourier Inversion:** *If $\widehat{f} \in \mathrm{L}^1(\mathbb{R})$, then f is continuous with $f(x) = \int_{\mathbb{R}} \widehat{f}(\omega) e^{2\pi i \omega x} d\omega$.*

(d) $\mathcal{F}$ is an injective operator.

1.5.2 Theorem and Definition. Plancherel Theorem If $f \in \mathrm{L}^1(\mathbb{R}) \cap \mathrm{L}^2(\mathbb{R})$, then $\widehat{f} \in \mathrm{L}^2(\mathbb{R})$, with $\|\widehat{f}\|_2 = \|f\|_2$. Hence $\mathcal{F} : \mathrm{L}^1 \cap \mathrm{L}^2(\mathbb{R}) \to \mathrm{L}^2(\mathbb{R})$ is isometric with respect to $\| \cdot \|_2$, and there exists a unique extension $\mathcal{F} : \mathrm{L}^2(\mathbb{R}) \to \mathrm{L}^2(\mathbb{R})$. This extension is necessarily isometric, and turns out to be unitary. Again, we write $\widehat{f} := \mathcal{F}(f)$. □

Theorem 1.5.3 *(a) For all $f \in \mathrm{L}^2(\mathbb{R})$: $f(x) = \widehat{\widehat{f}}(-x) = f(x)$ almost everywhere. In particular, if $\widehat{f} \in \mathrm{L}^1(\mathbb{R})$, then f is continuous with $f(x) = \int_{\mathbb{R}} \widehat{f}(\omega) e^{2\pi i \omega x} d\omega$.*

(b) (Sharpening of the Plancherel Theorem) Let $f \in \mathrm{L}^1(\mathbb{R})$. Then

$$f \in \mathrm{L}^2(\mathbb{R}) \Leftrightarrow \widehat{f} \in \mathrm{L}^2(\mathbb{R}) \quad .$$

1.5.4 Interpretation of the Fourier Transform An exponential function $e_\omega(x) = \cos(2\pi\omega x) + i \sin(2\pi\omega x)$ is interpreted as "pure frequency ω". The (informal) relation $\widehat{f}(\omega) = \langle f, e_\omega \rangle$ suggests interpreting $\widehat{f}(\omega)$ as "expansion coefficient" of f with respect to the pure frequency component, and the Fourier transform as a change of coordinates (in a similar way as outlined for ONBs in Remark 1.2.11). In this way the inversion formula

$$f(x) = \int_{\mathbb{R}} \widehat{f}(\omega) e_\omega(x) d\omega$$

can be read as the decomposition of f into its frequency components.

Moreover, the next section shows that the Fourier transform diagonalises translation and differentiation operators. □

Definition 1.5.5 Let $\omega, x, a \in \mathbb{R}$, with $a \neq 0$. We define

$$\begin{aligned} (T_x f)(y) &= f(y - x) && \text{(Translation operator)} \\ (D_a f)(y) &= |a|^{-1/2} f(a^{-1} y) && \text{(Dilation operator)} \\ (M_\omega f)(y) &= e^{-2\pi i \omega y} f(y) && \text{(Modulation operator)} \end{aligned}$$

All these operators are easily checked to be unitary on $\mathrm{L}^2(\mathbb{R})$. □

Dilation and translation have a simple geometrical meaning. T_x amounts to shifting a graph of a function by x, whereas D_a is a dilation of the graph by a along the x-axis, with a dilation by $|a|^{-1/2}$ along the y-axis: If $\mathrm{supp} f = [0, 1]$, then $\mathrm{supp}(D_a f) = [0, a]$ and $\mathrm{supp}(T_x f) = [x, x + 1]$.

1.5.6 Exercise

(a) Check the following relations on $\mathrm{L}^2(\mathbb{R})$

$$\begin{array}{rcl} \mathcal{F} \circ T_x = M_x \circ \mathcal{F} & , & \mathcal{F} \circ D_a = D_a \circ \mathcal{F} \\ \mathcal{F} \circ M_\omega = T_{-\omega} \circ \mathcal{F} & , & D_a \circ T_x = T_{ax} \circ D_a \\ T_x^* = T_{-x} \;\; , \;\; M_\omega^* = M_{-\omega} & , & D_a^* = D_{a^{-1}} \end{array}$$

(b) Verify for $f \in \mathrm{L}^2(\mathbb{R}) \cup \mathrm{L}^1(\mathbb{R})$:

$$\begin{array}{rcl} f(x) \in \mathbb{R} \text{ (a.e.)} & \Leftrightarrow & \widehat{f}(\omega) = \overline{\widehat{f}(-\omega)} \quad \text{(a.e.)} \\ f(x) = \overline{f(-x)} \text{ (a.e.)} & \Leftrightarrow & \widehat{f}(\omega) \in \mathbb{R} \quad \text{(a.e.)} \end{array}$$

□

One shortcoming of the Fourier transform, which will be discussed in more detail in the next section and serves as a motivation for the introduction of wavelets is the lack of localisation. There is quite a number of possible formulations of this lack of localisation. A particularly simple one is the following: If a function has compact support, its Fourier transform vanishes only on a discrete set.

Theorem 1.5.7 (Qualitative uncertainty principle) *Let $f \in \mathrm{L}^2(\mathbb{R})$ have compact support. Then its Fourier transform is analytic, in particular, if $f \neq 0$, $\widehat{f}$ vanishes only on a discrete set.*

As a consequence we see that if $g, h \in \mathrm{L}^1(\mathbb{R})$ only differ on a bounded interval, their Fourier transforms differ everywhere except on a discrete set. Hence a local change of the function g results in a global change of its Fourier transform.

1.6 Fourier Transforms and Differentiation

In this section we discuss how the decay behavior of the Fourier transform $\widehat{f}$ can be used to decide whether $f \in C^n(\mathbb{R})$ for some n. The results highlight both the usefulness and the restrictions of the Fourier transform: It is possible to derive smoothness properties from the decay of the Fourier transform, but only global ones. Due to the lack of localisation of the Fourier transform, functions violating the smoothness requirement in a single point cannot be distinguished from functions which are highly irregular.

We start with informal arguments: Since the exponential functions e_ω are smooth functions, with the oscillatory behavior controlled by the parameter ω, it seems natural to expect that reproducing a highly irregular function requires many exponentials with large frequencies. Thus the Fourier transform of such a function cannot be expected to be concentrated in small frequencies.

An alternative argument goes like this: In general, the differentiation operator lowers the regularity of a function; if $f \in C^n(\mathbb{R})$, then $f' \in C^{n-1}(\mathbb{R})$, but usually $f' \notin C^n(\mathbb{R})$. Hence differentiation increases irregularity. On the other hand, assuming we were allowed to differentiate under the integral of the Fourier inversion formula

$$f(t) = \int_{\mathbb{R}} \widehat{f}(\omega) e^{2\pi i \omega t} d\omega$$

we obtain

$$f'(t) = \int_{\mathbb{R}} \widehat{f}(\omega)(2\pi i\omega)e^{2\pi i\omega t}d\omega \quad .$$

Hence we have "derived" $\widehat{f'}(\omega) = \widehat{f}(\omega)(2\pi i\omega)$. Thus differentiation results in a multiplication with the frequency variable; small frequencies are attenuated while large frequencies become more important. The first result of this section contains a rigourous formulation of this observation:

Theorem 1.6.1 *(a) Let $f \in \mathrm{L}^1(\mathbb{R}) \cap C^n(\mathbb{R})$, with $f^{(i)} \in \mathrm{L}^1(\mathbb{R})$ for $1 \le i \le n$. Then*

$$\mathcal{F}(f^{(n)})(\omega) = (2\pi i\omega)^n \widehat{f}(\omega) \quad .$$

In particular $|\widehat{f}(\omega)|\ |\omega|^n \to 0$ for $\omega \to 0$.

(b) Conversely, if $f \in \mathrm{L}^1(\mathbb{R})$ with $|\widehat{f}(\omega)| \le C|\omega|^{-(n+\alpha)}$ for some $\alpha > 1$, then $f \in C^n(\mathbb{R})$.

Remark 1.6.2 The last theorem contains conditions on the decay of $\widehat{f}$ for f to be C^n, namely

Necessary:	$\lvert\widehat{f}(\omega)\rvert\ \lvert\omega\rvert^n \to 0$
Sufficient:	$\lvert\widehat{f}(\omega)\rvert\ \lvert\omega\rvert^n \le \lvert\omega\rvert^{-\alpha} \quad (\alpha > 1)$

Note that there is a gap between the necessary and sufficient conditions. In fact, there are functions sitting in this gap; see below.

On the other hand, the sufficient condition shows that if f is not C^n, **even at a single point in** $\mathbb{R}$, there does not exist $\alpha > 1$ with a constant $C > 0$ such that $|\widehat{f}(\omega)| \le C|\omega|^{\alpha+n}$. In other words, a single point of lower regularity influences the global decay behaviour of $\widehat{f}$. □

Example 1.6.3 (a) Consider $f(x) = \chi_{[-S,S]}(x)$ for some $S > 0$. Then f is piecewise C^∞, with singularities $-S, S$. Straightforward computation yields

$$\widehat{f}(\omega) = \frac{\sin(2\pi\omega S)}{\pi\omega} \quad .$$

In particular, we obtain $|\widehat{f}(\omega)| \le C(1 + |\omega|)^{-1}$. Moreover, this decay order is optimal, since otherwise $\widehat{f}$ were integrable and hence f continuous by the Fourier inversion Theorem 1.5.1.

(b) Let $f(x) = e^{-|x|}$. Then f is continuous and C^∞ in $\mathbb{R} \setminus \{0\}$. A straightforward computation establishes

$$\widehat{f}(\omega) = \frac{1}{1 + 4\pi^2\omega^2} \quad .$$

Hence $|\widehat{f}(\omega)| \le C(1 + |\omega|)^{-2}$, but $f \notin C^1$. This shows that the sufficient decay condition in Remark 1.6.2 cannot be extended to $\alpha = 1$. □

Remark 1.6.4 The examples can also be read as a sharpening of the remark following Theorem 1.5.7. If $f \in \mathrm{L}^1(\mathbb{R})$ is smooth except for a few isolated singularities, one can easily construct a smooth $g \in \mathrm{L}^1(\mathbb{R})$ which only differs from f in a neighborhood of the singularities. $\widehat{g}$ decays significantly faster than $\widehat{f}$, i.e., the decay behavior of f is determined by the difference $f - g$. This difference only depends on the behavior of f close to the singularities. Hence the Fourier transform is not well adapted to the treatment of piecewise smooth functions. □

1.6.5 Exercise Let $f \in \mathrm{L}^1(\mathbb{R})$. Assume that

$$\int_{\mathbb{R}} |t|^n |f(t)| dt < \infty$$

for some $n \geq 1$, then $\widehat{f} \in C^n$ with $\left(\widehat{f}\right)^{(n)} = (-2\pi i)^n (t^n f)^\wedge$. □

1.6.6 Exercise Compute the Fourier transform of $f(x) = e^{-x^2/2}$.
(Hint: First derive the differential equation

$$\left(\widehat{f}\right)'(\omega) = -4\pi^2 \omega \widehat{f}(\omega)$$

and solve it. The correct normalisation is obtained by

$$\widehat{f}(0) = \int_{\mathbb{R}} e^{-x^2/2} dx = \sqrt{2\pi} \quad .$$

A proof for the second equality is not required.) □

1.6.7 Exercise A function $f : \mathbb{R} \to \mathbb{C}$ is called **Schwartz-function** when $f \in C^\infty(\mathbb{R})$ with $|x^m f^{(n)}(x)| \leq C_{m,n}(f)(f)$, for all $m, n \in \mathbb{N}$ and $x \in \mathbb{R}$, and constants $C_{m,n}(f)$. The space of Schwartz-functions is denoted by $\mathcal{S}(\mathbb{R})$. Prove that

$$f \in \mathcal{S}(\mathbb{R}) \Leftrightarrow \widehat{f} \in \mathcal{S}(\mathbb{R}) \quad .$$

□

Chapter 2

The continuous wavelet transform

2.1 Admissibility, Inversion Formula

Definition 2.1.1 $\psi \in \mathrm{L}^2(\mathbb{R})$ is called **admissible** (or **(mother) wavelet** if $\psi \neq 0$ and

$$c_\psi = \int_{\mathbb{R}} \frac{|\widehat{\psi}(\omega)|^2}{|\omega|} d\omega < \infty \quad . \tag{2.1}$$

□

Remark 2.1.2 Let $\psi \in \mathrm{L}^2(\mathbb{R})$ with $\int_{\mathbb{R}} |t||\psi(t)|dt < \infty$. Then $\psi \in \mathrm{L}^1(\mathbb{R})$ and

$$\psi \text{ is admissible } \Leftrightarrow \widehat{\psi}(0) = \int_{\mathbb{R}} \psi(t)dt = 0 \tag{2.2}$$

Proof.

$$\begin{aligned} \int_{\mathbb{R}} |\psi(t)|dt &= \left(\int_{-1}^{1} + \int_{|t|>1} \right) |\psi(t)|dt \\ &\leq \int_{-1}^{1} |\psi(t)|dt + \int_{|t|>1} |t||\psi(t)|dt \quad . \end{aligned}$$

The first term is finite since the restriction of ψ to $[-1,1]$ is in $\mathrm{L}^2([-1,1]) \subset \mathrm{L}^1([-1,1])$, and the second term is finite by assumption. Thus $\psi \in \mathrm{L}^1(\mathbb{R})$. Moreover, $\widehat{\psi}$ is C^1, by 1.6.5, hence there exists a constant c such that for all $|\omega| < 1$ the inequality $|\widehat{\psi}(0) - \widehat{\psi}(\omega)| \leq c|\omega|$ holds. Thus

$$\int_{-1}^{1} \frac{|\widehat{\psi}(\omega)|^2}{|\omega|} d\omega \geq \int_{-1}^{1} \frac{|\widehat{\psi}(0)|^2}{|\omega|} - c^2|\omega| d\omega \quad ,$$

and the right hand side is finite only for $\widehat{\psi}(0) = 0$. This shows $\Leftarrow$ of (2.2), whereas similar considerations establish the other direction. □

Definition 2.1.3 Let $f \in \mathrm{L}^2(\mathbb{R})$, $\psi \in \mathrm{L}^2(\mathbb{R})$ admissible. The **continuous wavelet transform of f with respect to ψ** (**CWT**, for short) is the function $W_\psi f : \mathbb{R} \times \mathbb{R}^* \to \mathbb{C}$, defined by

$$\begin{aligned} (W_\psi f)(b,a) &= \langle f, T_b D_a \psi \rangle \\ &= \int_{\mathbb{R}} f(x) \frac{1}{\sqrt{|a|}} \overline{\psi\left(\frac{x-b}{a}\right)} dx \quad . \end{aligned}$$

□

Remark 2.1.4 Let $f, g \in \mathrm{L}^2(\mathbb{R})$ (admissible or not). The function $(W_g f)(b,a) = \langle f, T_b D_a g \rangle$ has the following properties: It is continuous and vanishes at infinity on $\mathbb{R} \times \mathbb{R}^*$. The latter means the following: Given any sequence of points $((a_n, b_n))_{n \in \mathbb{N}} \subset \mathbb{R} \times \mathbb{R}^*$ without accumulation point in $\mathbb{R} \times \mathbb{R}^*$,

$$W_g f(b_n, a_n) \to 0, \quad \text{as } n \to \infty \quad .$$

Proof. Exercise. (Hint: First plug in characteristic functions of compact sets. The Cauchy-Schwarz inequality allows to extend the results to arbitrary elements of $\mathrm{L}^2(\mathbb{R})$.) □

2.1.5 Examples

(a) **Haar wavelet:** $\psi = \chi_{[0,1/2[} - \chi_{[1/2,1]}$. ψ is admissible by (2.2).

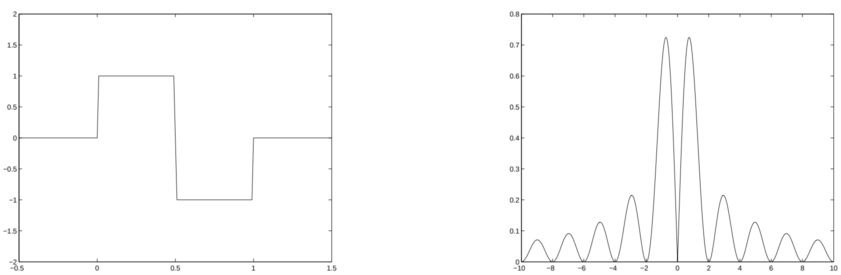

Figure 2.1: The Haar wavelet and its Fourier transform (only the absolute value)

(b) **Mexican Hat wavelet:** Define $\psi(x) = -\frac{d^2}{d^2y}(e^{-y^2/2})(x) = (1-x^2)e^{-x^2/2}$. By 1.6.1 we compute the Fourier transform as

$$\widehat{\psi}(\omega) = (2\pi i \omega)^2 \mathcal{F}(e^{-\cdot^2/2})(\omega) = -4\pi^2\omega^2\sqrt{2\pi}e^{-2\pi^2\omega^2} \quad .$$

In particular $\widehat{\psi}(0) = 0$, i.e. ψ is admissible by (2.2).

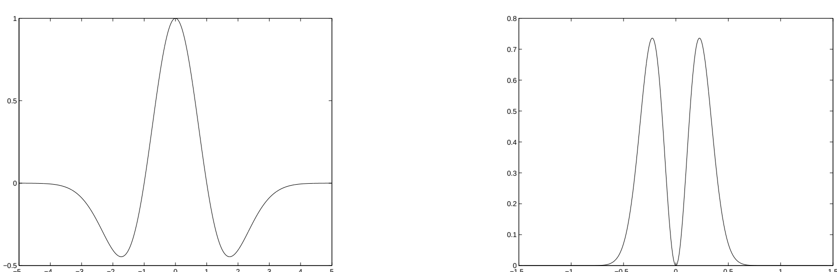

Figure 2.2: The mexican hat wavelet and its Fourier transform

(c) **Shannon wavelet:** Define ψ by $\widehat{\psi} = \chi_{[-1,-1/2]} + \chi_{[1/2,1]}$. The admissibility condition (2.1) is easily checked. Using 1.5.3 (a) and 1.6.3 (a), the inverse Fourier transform of ψ is computed as

$$\psi(x) = \frac{\sin(\pi x)}{\pi x}(2\cos(\pi x) - 1) \quad .$$

Unlike the previous two examples, the Shannon wavelets decays rather slowly ($|\psi(x)| \sim |x|^{-1}$). In fact, $\psi \notin \mathrm{L}^1(\mathbb{R})$, since $\widehat{\psi}$ is not continuous. □

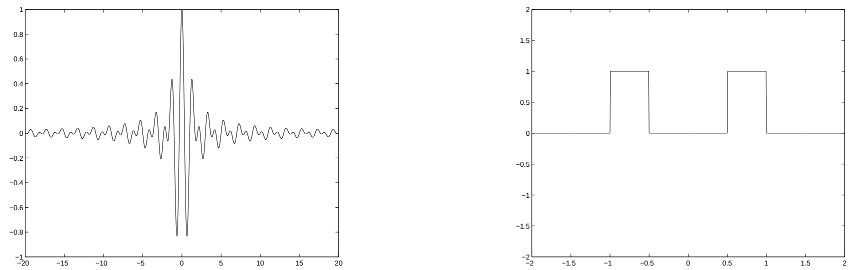

Figure 2.3: The Shannon wavelet and its Fourier transform

Theorem 2.1.6 *Let* $f \in \mathrm{L}^2(\mathbb{R})$, $\psi \in \mathrm{L}^2(\mathbb{R})$ *admissible. Then* $W_\psi f \in \mathrm{L}^2(\mathbb{R} \times \mathbb{R}^*, |a|^{-2} db da)$, *with*

$$\int_{\mathbb{R}^*} \int_{\mathbb{R}} |W_\psi f(b,a)|^2 \, db \frac{da}{|a|^2} = c_\psi \|f\|_2^2 \quad . \tag{2.3}$$

In other words: The operator $f \mapsto c_\psi^{-1/2} W_\psi f$ *is an isometry* $\mathrm{L}^2(\mathbb{R}) \to \mathrm{L}^2(\mathbb{R} \times \mathbb{R}^*, |a|^{-2} db da)$.

Proof. We compute

$$\begin{aligned}
\int_{\mathbb{R}^*} \int_{\mathbb{R}} |W_\psi f(b,a)|^2 db \frac{da}{|a|^2} &= \int_{\mathbb{R}^*} \int_{\mathbb{R}} |\langle f, T_b D_a \psi|^2 db \frac{da}{|a|^2} \\
&= \int_{\mathbb{R}^*} \int_{\mathbb{R}} |\langle \widehat{f}, M_{-b} D_{a^{-1}} \widehat{\psi} \rangle|^2 db \frac{da}{|a|^2} \\
&= \int_{\mathbb{R}^*} \int_{\mathbb{R}} \left| \int_{\mathbb{R}} \widehat{f}(\omega) |a|^{1/2} \overline{\widehat{\psi}(a\omega)} e^{2\pi i b \omega} d\omega \right|^2 db \frac{da}{|a|^2} \\
&= \int_{\mathbb{R}^*} \int_{\mathbb{R}} |\mathcal{F}(h_a)(-b)|^2 \, db \frac{da}{|a|^2} \quad ,
\end{aligned}$$

where $h_a(\omega) = \widehat{f}(\omega)|a|^{1/2}\overline{\widehat{\psi}(a\omega)}$, which for fixed a is an L^1-function. Thus applying 1.5.3 (b) we obtain that

$$\begin{aligned}
\int_{\mathbb{R}^*} \int_{\mathbb{R}} |\mathcal{F}(h_a)(-b)|^2 \, db \frac{da}{|a|^2} &= \int_{\mathbb{R}^*} \int_{\mathbb{R}} \left| \widehat{f}(\omega) |a|^{1/2} \overline{\widehat{\psi}(a\omega)} \right|^2 d\omega \frac{da}{|a|^2} \\
&= \int_{\mathbb{R}^*} \int_{\mathbb{R}} \left| \widehat{f}(\omega) \overline{\widehat{\psi}(a\omega)} \right|^2 d\omega \frac{da}{|a|} \\
&= \int_{\mathbb{R}} |\widehat{f}(\omega)|^2 \int_{\mathbb{R}^*} |\widehat{\psi}(a\omega)|^2 \frac{da}{|a|} \, d\omega \\
&= \int_{\mathbb{R}} |\widehat{f}(\omega)|^2 \int_{\mathbb{R}^*} |\widehat{\psi}(a)|^2 \frac{da}{|a|} \, d\omega \\
&= c_\psi \|f\|_2^2 \quad .
\end{aligned}$$

□

Given arbitrary nonzero f, the proof shows that the property $W_\psi f \in \mathrm{L}^2(\mathbb{R} \times \mathbb{R}^*, |a|^{-2} db da)$ only depends on the choice of ψ.

2.1.7 Inversion Formulae

(a) The isometry property of the operator $\mathcal{W}_\psi$ is often stated in the form of **inversion**

formulae, i.e., via the equations

$$f = \frac{1}{c_\psi} \int_{\mathbb{R}} \int_{\mathbb{R}^*} W_\psi f(b,a)(T_b D_a \psi) \frac{da}{|a|^2} db \quad (2.4)$$

$$= \frac{1}{c_\psi} \langle f, T_b D_a \psi \rangle (T_b D_a \psi) \frac{da}{|a|^2} db \quad . \quad (2.5)$$

Here the right hand side converges only in the weak sense, which is just a restatement of the isometry property. In particular, it should not be read pointwise (we will encounter pointwise inversion formulae below).
(b) Let $f, \psi \in \mathrm{L}^2(\mathbb{R})$, with ψ admissible. If f or ψ are real-valued, the upper half plane is sufficient for reconstruction, i.e. instead of (2.4) we have

$$f = \frac{2}{c_\psi} \int_{\mathbb{R}} \int_{\mathbb{R}^+} W_\psi f(b,a)(T_b D_a \psi) \frac{da}{|a|^2} db \quad . \quad (2.6)$$

Proof. Let

$$I_+ = \int_{\mathbb{R}} \int_{\mathbb{R}^+} |W_\psi f(b,a)|^2 \frac{da}{|a|^2} db$$

and I_- defined analogously for the lower half plane. The same computation as in the proof of (2.3) yields that

$$I_{+/-} = \int_{\mathbb{R}} \int_{\mathbb{R}^{+/-}} |h_a(\omega)|^2 d\omega \frac{da}{|a|^2} \quad ,$$

where $h_a(\omega) = \widehat{f}(\omega)|a|^{1/2}\overline{\widehat{\psi}(a\omega)}$. Supposing $\psi = \overline{\psi}$ we would obtain $\overline{\widehat{\psi}(a\omega)} = \widehat{\psi}(-a\omega)$, and thus $|h_a(\omega)| = |h_{-a}(\omega)|$. If $f = \overline{f}$, then

$$|h_a(\omega)| = |\widehat{f}(\omega)|a|^{1/2}\widehat{\psi}(a\omega)| = |\overline{\widehat{f}(-\omega)}|a|^{1/2}\widehat{\psi}(a\omega)| = |h_{-a}(-\omega)| \quad .$$

In both case $I_+ = I_-$, and since $I_+ + I_- = c_\psi \|f\|_2^2$, $I_{+/-} = \frac{c_\psi}{2}\|f\|_2^2$ follows. □

2.1.8 Interpretation (I)

The wavelet ψ is declared a detail of scale 1 at position 0. Then $T_b D_a \psi$ is a detail of scale a at position b. The inversion formula

$$f = \frac{1}{c_\psi} \int_{\mathbb{R}} \int_{\mathbb{R}^*} W_\psi f(b,a)(T_b D_a \psi) \frac{da}{|a|^2} db$$

should be read as a decomposition of f into its details at varying scales and positions, where $W_\psi f(b,a) = \langle f, T_b D_a \psi \rangle$ is precisely the "expansion coefficient" one expects from the context of ONB's. (Compare also to Fourier integrals.) However, the analogy to ONB's or Fourier integrals is only limited. In particular, the operator W_ψ is never onto all of $\mathrm{L}^2(\mathbb{R} \times \mathbb{R}^*, |a|^{-2} dbda)$. That means, we cannot prescribe arbitrary coefficients, i.e., any function $W \in \mathrm{L}^2(\mathbb{R} \times \mathbb{R}^*, |a|^{-2} dbda)$, and expect that it is the wavelet transform of an $f \in \mathrm{L}^2(\mathbb{R})$. Indeed, by 2.1.4 W would at least have to be continuous. A more explicit description of the image space $W_\psi(\mathrm{L}^2(\mathbb{R}))$ is contained in the following proposition. □

Proposition 2.1.9 *Let $\psi \in \mathrm{L}^2(\mathbb{R})$ be admissible. Define for $F \in \mathrm{L}^2(\mathbb{R} \times \mathbb{R}^*, |a|^{-2} dbda)$ the function*

$$(PF)(b,a) = \frac{1}{c_\psi} \int_{\mathbb{R}} \int_{\mathbb{R}}^* F(b',a') \langle T_{b'} D_{a'} \psi, T_b D_a \psi \rangle \frac{da}{|a|^2} db \quad .$$

This converges absolutely for all $(b,a) \in \mathbb{R} \times \mathbb{R}^*$, and $F \mapsto PF$ is the projection onto the closed subspace $W_\psi(\mathrm{L}^2(\mathbb{R}))$. In particular, $F \in W_\psi(\mathrm{L}^2(\mathbb{R}))$ iff it satisfies the **reproducing kernel relation**

$$F(b,a) = \frac{1}{c_\psi} \int_{\mathbb{R}} \int_{\mathbb{R}^*} F(b',a') \langle T_{b'} D_{a'} \psi, T_b D_a \psi \rangle \frac{da}{|a|^2} db \quad . \tag{2.7}$$

Proof. The key observation is that the "kernel function"

$$(b',a',b,a) \mapsto \langle T_{b'} D_{a'} \psi, T_b D_a \psi \rangle$$

is a wavelet transform itself, namely for fixed (b,a) we have

$$\langle T_{b'} D_{a'} \psi, T_b D_a \psi \rangle = \overline{[W_\psi(T_b D_a \psi)]\,(b',a')} \ ,$$

and thus

$$(PF)(b,a) = \frac{1}{c_\psi} \langle F, W_\psi(T_b D_a \psi) \rangle_{\mathrm{L}^2(\mathbb{R}\times\mathbb{R}^*,|a|^{-2}dbda)}$$

is pointwise well-defined for all (b,a).

Now let us show that P is the orthogonal projection onto $W_\psi(\mathrm{L}^2(\mathbb{R})) \subset \mathrm{L}^2(\mathbb{R}\times\mathbb{R}^*, |a|^{-2}dbda)$; note that it is a closed subspace. Write $F = F_1 + F_2$ with $F_2 \perp W_\psi(\mathrm{L}^2(\mathbb{R}))$, $F_1 = W_\psi g$ for some $g \in \mathrm{L}^2(\mathbb{R})$. Then

$$\begin{aligned}
(PF)(b,a) &= \frac{1}{c_\psi} \langle F, W_\psi(T_b D_a \psi) \rangle_{\mathrm{L}^2(\mathbb{R}\times\mathbb{R}^*,|a|^{-2}dbda)} \\
&= \frac{1}{c_\psi} \langle W_\psi g + F_2, W_\psi(T_b D_a \psi) \rangle_{\mathrm{L}^2(\mathbb{R}\times\mathbb{R}^*,|a|^{-2}dbda)} \\
&= \frac{1}{c_\psi} \langle W_\psi g, W_\psi(T_b D_a \psi) \rangle_{\mathrm{L}^2(\mathbb{R}\times\mathbb{R}^*,|a|^{-2}dbda)} \\
&= \langle g, T_b D_a \psi \rangle_{\mathrm{L}^2(\mathbb{R})} = F_1 \quad ,
\end{aligned}$$

where we have used the fact that isometries respect scalar products (1.2.7(a)). □

2.1.10 Interpretation of CWT (II) Let $\psi \in \mathrm{L}^2(\mathbb{R})$ be admissible and $f \in \mathrm{L}^2(\mathbb{R})$. In this remark we want to look at restrictions of $W_\psi f$ to horizontal lines, and their information content. Recycling the argument from the proof of (2.3), we saw that

$$W_\psi f(b,a) = \mathcal{F}(h_a)(-b) = \mathcal{F}^{-1}(h_a)(b) \quad ,$$

where $h_a(\omega) = \widehat{f}(\omega)|a|^{1/2}\overline{\widehat{\psi}(a\omega)}$. Since $W_\psi f \in \mathrm{L}^2(\mathbb{R} \times \mathbb{R}^*, |a|^{-2}dbda)$, $W_\psi f(\cdot,a) \in \mathrm{L}^2(\mathbb{R})$ for almost all $a \in \mathbb{R}^*$, with Fourier transform

$$\omega \mapsto \widehat{f}(\omega)|a|^{1/2}\overline{\widehat{\psi}(a\omega)} \quad .$$

In order to better appreciate this formula, it is useful to take a second look at the Fourier transforms of the candidate wavelets we have seen so far: Taking the Shannon wavelet ψ, we obtain $D_{a^{-1}}\widehat{\psi} = \sqrt{|a|}\left(\chi_{[-1/a,-1/2a]} + \chi_{[1/2a,1/a]}\right)$. It follows that the Fourier transform of $W_\psi f(\cdot,a)$ is precisely the restriction of $\widehat{f}$ to the "frequency band"

$$\Omega_a = \{\omega : 1/2a \le |\omega| \le 1/a\} \quad .$$

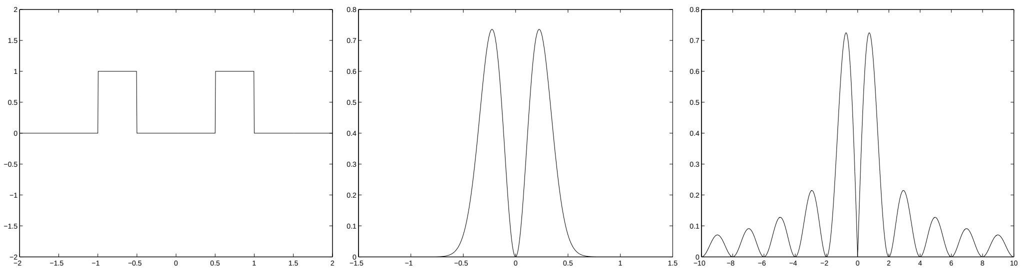

Figure 2.4: Fourier transforms of Shannon, Mexican Hat and Haar wavelets

On the other hand, the (however poor) decay behaviour of the Shannon wavelet implies that $\langle f, T_b D_a \psi \rangle$ contains information about f near b, i.e., we have a (however loose) interpretation

$$W_\psi f(b,a) \approx \text{ coefficient corresponding to frequency band } \Omega_a \text{ and position } b \quad .$$

If we are interested in strengthening the dependence of $W_\psi f(b,a)$ on the behaviour of f only in a neighborhood of b, a wavelet with compact support is a good choice, such as the Haar wavelet. However, a look at the Fourier transform of this wavelet shows that we have paid for this in terms of Fourier transform localisation.

However, there are many possible compromises, as e.g. the mexican hat wavelet. On the Fourier transform side, the sharp "boxcar" shape of the shannon wavelet is then replaced by two bumps, with sharp decrease towards $\pm\infty$. This is still superior to the shape of the Fourier transform of the Haar wavelet. The dilation parameter moves the bumps across the halfaxes, just as with the boxcar before. On the other hand, the loss of localisation on the Fourier transform side buys a much better decay of the wavelet ψ. Hence the wavelet coefficient $W_\psi(b,a)f$ depends much more on the behaviour of f in a small neighborhood of b than this is the case for the Shannon wavelet.

Summarising, handwaving arguments have provided us with an interpretation

$$W_\psi f(b,a) \approx \text{ frequencies from } \Omega_a, \text{ contained in } S_{b,a} \tag{2.8}$$

where Ω_a is an interval with center and length proportional to a (and its negative) and S is an interval with center b and length proportional to $1/a$. How well this interpretation works, depends on the choice of ψ: The sharper the localisation (measured by some decay of ψ and $\widehat{\psi}$), the better.

Hence a wavelet with compact supported both in time and frequency would be optimal, since it would allow to make the interpretation precise. But by 1.5.7 such a function does not exist. $\square$

2.1.11 Exercise (Convolution) (a) Show for $f, g \in \mathrm{L}^1(\mathbb{R})$, that the function

$$f * g(x) := \int_{\mathbb{R}} f(t) g(x-t) dt$$

is almost everywhere well-defined (i.e., the right-hand side is almost everywhere defined), and is in $\mathrm{L}^1(\mathbb{R})$, with $\|f * g\|_1 \le \|f\|_1 \|g\|_1$.
(b) Prove the **convolution theorem**, i.e., $\widehat{f * g}(\omega) = f(\omega) g(\omega)$.
(c) Now let $f, g \in \mathrm{L}^2(\mathbb{R})$, then

$$f * g(x) := \int_{\mathbb{R}} f(t) g(x-t) dt$$

converges absolutely for all $x \in \mathbb{R}$, by the Cauchy-Schwarz inequality. Prove that $f*g \in \mathrm{L}^2(\mathbb{R})$ iff $\widehat{f} \cdot \widehat{g} \in \mathrm{L}^2(\mathbb{R})$ (Hint: Use 1.5.3). Show that in this case, $\widehat{f*g} = \widehat{f} \cdot \widehat{g}$.
(d) Show that $W_\psi f(b,a) = f * D_a(\psi^*)$, where $\psi^*(x) = \overline{\psi(-x)}$. □

Remarks 2.1.12 (Linear Filters)
(a) Let $m : \mathbb{R} \to \mathbb{C}$ be a bounded measurable function. Then the associated **multiplication operator** $S_m : \mathrm{L}^2(\mathbb{R}) \to \mathrm{L}^2(\mathbb{R})$, $S_m(f)(x) = m(x)f(x)$ is a bounded linear operator. An operator $R : \mathrm{L}^2(\mathbb{R}) \to \mathrm{L}^2(\mathbb{R})$ is called **(linear) filter** if there exists $m \in \mathrm{L}^\infty(\mathbb{R})$ such that $R = \mathcal{F}^{-1} \circ S_m \circ \mathcal{F}$, i.e. $\forall f \in \mathrm{L}^2(\mathbb{R})$,

$$(Rf)^\wedge(\omega) = \widehat{f}(\omega)m(\omega) \quad .$$

The following filters best convey the idea:

- Ideal lowpass: $m_l = \chi_{[-T,T]}$.
- Ideal bandpass: $m_b = \chi_{[-\omega_2,\omega_1]} + \chi_{[\omega_1,\omega_2]}$.
- Ideal bandpass: $m_h = 1 - \chi_{[-T,T]}$.

These examples illustrate the "filter" language: Each (low-/band-/high-) pass filter picks out those frequencies of f which are in the prescribed range, and kills the others.

Another class of filters is given by convolution operators, $f \mapsto f * g$, where g is some function with bounded Fourier transform (e.g., $g \in \mathrm{L}^1(\mathbb{R})$). Here the frequencies are usually not just kept or killed, but rather attenuated or enhanced (depending on whether $|\widehat{g}(\omega)|$ is small or large). An exception is the Shannon wavelet, which akts as a a bandpass filter.

Returning to ideal bandpass filters, if we pick the examples m_l and m_h as above, using the same T, we obtain a first simple example of a **filterbank**:

$$f \mapsto (R_{m_l}f, R_{m_h}f)$$

decomposes f into its low-frequency and high-frequency parts. The particular choice of m_l and m_h in addition allow to reconstruct f, simply letting $f = R_{m_l}f + R_{m_h}f$. In fact, one can easily see that any decomposition

$$\mathbb{R} = \dot{\bigcup_{i\in\mathbb{Z}}} \Omega_i$$

of the frequency axis into intervals gives rise, via the corresponding bandpass filters, to a filterbank

$$f \mapsto (R_i f)_{i\in\mathbb{Z}} \quad ,$$

with the property that $f = \sum_{i\in\mathbb{Z}} R_i f$. To see the connection between this notion and CWT, we let $\Omega_i = [-2^{-i}, -2^{-1+1}[\cup]2^{-i+1}, 2^{-i}]$. Then we see that $R_i f(b) = W_\psi f(b, 2^i)$, where ψ denotes the Shannon wavelet. Thus the Shannon CWT can be seen as a (highly redundant) filterbank. This observation generalises by the previous exercise about convolutions to arbitrary wavelets:

Whenever ψ is admissible, the associated CWT acts as a continuously indexed filterbank, i.e., a family of convolution operators

$$W_\psi : f \mapsto (f * D_a\psi^*)_{a\in\mathbb{R}^*} \quad .$$

This view of CWT as a filterbank can be best understood by studying a similar transform, called ”dyadic wavelet transform”. This will be done in an exercise below.

Let us give an illustration of the filterbank terminology. Consider the function $f : [0, 1] \to \mathbb{R}$ given by

$$f(x) = \sin(96\pi x) + \sin(64\pi^2 x^2) \quad .$$

f is the sum of two components, the first one of constant frequency, while in the second one corresponds the frequency increases with time. Visual inspection of the plot tells little more

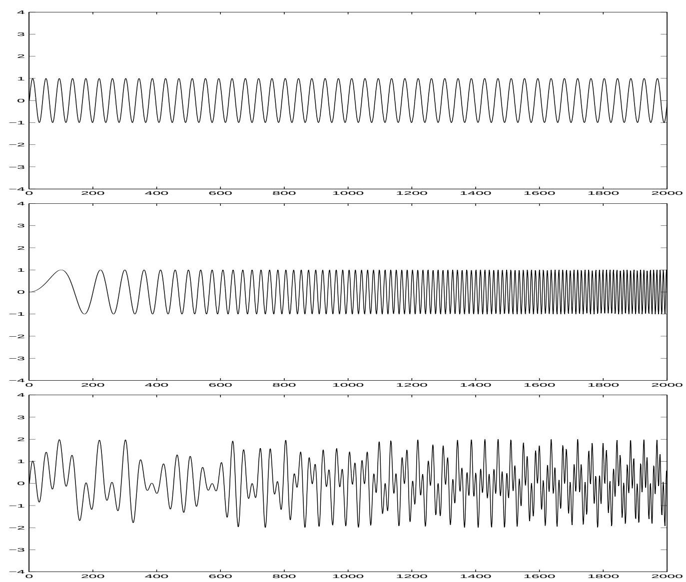

Figure 2.5: A signal with one component of constant frequency and one component of varying frequency

than that the function becomes more “wiggly” with time. On the other hand, looking at the continuous wavelet transform with the Shannon wavelet as analysing wavelet we clearly see the two components as two strips: One of constant scale, and the other one with decreasing scale (and thus increasing frequency).

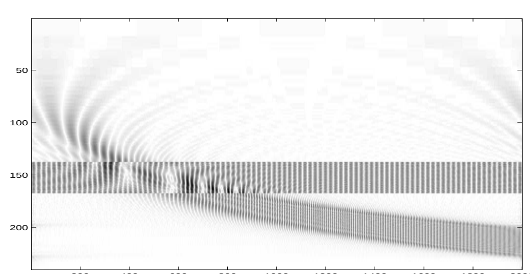

Figure 2.6: The wavelet coefficients for the Shannon wavelet.

Comparison for other wavelets shows similar behaviour, though somewhat blurred due to less sharp concentration in frequency.

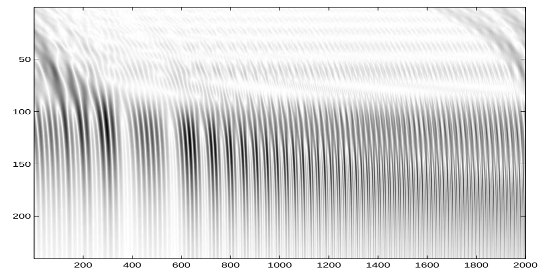

Figure 2.7: The wavelet coefficients for the Haar wavelet. The Haar wavelet is too poorly concentrated on the frequency side to give a clear picture.

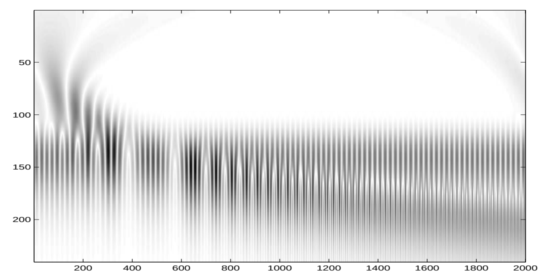

Figure 2.8: The wavelet coefficients for the Mexican Hat wavelet, which does a slightly better job than the Haar wavelet.

However, the best results are obtained for the so-called *Morlet wavelet*, which is defined essentially as a gaussian, shifted on the frequency side:

$$\psi(x) = e^{-x^2/2} e^{i\omega_0 x} + \kappa(x) \quad .$$

Here κ is a (negligible) correction term ensuring $\widehat{\psi}(x) = 0$. The quick decay in both time and frequency yield the following sharp picture:

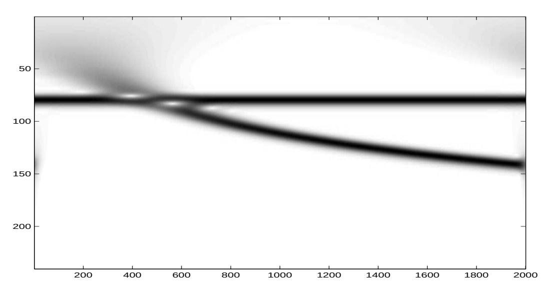

Figure 2.9: The wavelet coefficients for the Morlet wavelet

These examples yield a first illustration of the properties of continuous wavelet transforms. They also show the advantage of having a large arsenal of wavelets. Signals of the type (2.9) are called “chirps”. They arise in signal processing, for instance in audio signals. However, we just used them for illustration purposes; our main signal class are piecewise smooth functions.

This discussion should also remind the reader of Remark 1.2.11: Picking a wavelet, we introduce a coordinate system in $L^2(\mathbb{R})$ corresponding to the family of scaled and shifted copies of the wavelet. And choosing the right wavelet yields a particularly compact representation of "interesting" signals such as chirps.

(b) An extreme case of a linear filter is a translation operator. As a general rule, linear filters commute (this is obvious on the Fourier side). There is a converse: If an operator T commutes with all translations, then it is necessarily a linear filter. □

Lemma 2.1.13 *(a) If f, g have compact support, so does $f * g$.*
(b) Assume that $f, g \in \mathrm{L}^2(\mathbb{R})$ or $f, g \in \mathrm{L}^1(\mathbb{R})$. Then $f * g = g * f$.

2.1.14 Exercise: Dyadic Wavelet Transform For given $f, \psi \in \mathrm{L}^2(\mathbb{R})$ we let the **dyadic wavelet transform of f with respect to ψ** be the function living on $\mathbb{R} \times \mathbb{Z}$, defined as

$$\widetilde{W_\psi} f(b, n) = W_\psi f(b, 2^n) = \langle f, T_b D_{2^n} \psi \rangle \quad .$$

ψ is called **dyadically admissible** if for all $f \in \mathrm{L}^2(\mathbb{R})$ the normequality

$$\|f\|_2^2 = \sum_{n \in \mathbb{Z}} 2^{-n} \int_{\mathbb{R}} |\widetilde{W_\psi} f(b, n)|^2 db \tag{2.9}$$

holds.

(a) Show that ψ is dyadically admissible iff

$$\sum_{n \in \mathbb{Z}} |\widehat{\psi}(2^n \omega)|^2 = 1$$

holds, for almost all $\omega \in \mathbb{R}$. Show that it is enough to check ω with $1/2 < |\omega| \leq 1$.

(b) Conclude that if ψ is dyadically admissible, the mappings $g \mapsto g * (D_{2^n} \psi)$ and $g \mapsto g * (D_{2^n} \psi^*)$ define bounded convolution operators on $\mathrm{L}^2(\mathbb{R})$.

(c) Show, for dyadically admissible ψ the reconstruction formulae

$$\begin{aligned} f &= \sum_{n \in \mathbb{Z}} 2^{-n} (f * (D_{2^n} \psi^*)) * (D_{2^n} \psi) \\ &= \sum_{n \in \mathbb{Z}} 2^{-n} (\widetilde{W_\psi} f(\cdot, n)) * (D_{2^n} \psi) \end{aligned}$$

with convergence in $\| \cdot \|_2$.

(d) Show that if ψ is dyadically admissible, it is also admissible in the sense of (2.1).

□

2.1.15 The affine group The parameter space $\mathbb{R} \times \mathbb{R}^*$ of the CWT may be identified canonically with the group of affine mappings $\mathbb{R} \to \mathbb{R}$. Indeed, for $(b, a) \in \mathbb{R} \times \mathbb{R}^*$ we define a map $r_{b,a}$ via $r_{b,a}$, $r_{b,a}(x) = ax + b$. Then concatenation of two such mappings corresponds to the multiplication

$$(b, a) \circ (b', a') = (b + ab', aa') \quad ,$$

which endows $G = \mathbb{R} \times \mathbb{R}'$ with a group multiplication. The inversion is easily shown to be

$$(b, a)^{-1} = (-a^{-1} b, a^{-1}) \quad .$$

Moreover, it is a locally compact topological group, which means that the group operations are continuous (and $\mathbb{R} \times \mathbb{R}'$ with the product topology is locally compact). It turns out that some of the conditions and definitions in this section can be interpreted in the context of this group. It turns out, that analogous constructions can be used to construct CWT in many other settings, e.g., in higher dimensions.

(a) The mapping $\pi : (b, a) \mapsto T_b D_a$ is a continuous group homomorphism $G \to \mathcal{U}(L^2(\mathbb{R}))$, where the right-hand side is the group of unitary operators on $\mathrm{L}^2(\mathbb{R})$. Here continuity refers to the so-called strong operator topology, meaning that $g_n \to g$ in G entails $\pi(g_n)f \to \pi(g)f$, for all $f \in \mathrm{L}^2(\mathbb{R})$.

(b) The measure $d\mu(b, a) = db\frac{da}{|a|^2}$ is leftinvariant, meaning that for every set $A \subset G$ and every $g \in G$, the set $gA = \{gh : h \in A\}$ has the same measure: $\mu(gA) = \mu(A)$. An alternative formulation is the following: Given $F \in \mathrm{L}^2(\mathbb{R} \times \mathbb{R}^*, |a|^{-2}dbda)$, the **left translate**, defined to be the function $F_{(b_0,a_0)}(b, a) = F((b_0, a_0)^{-1}(b, a))$ is in $\mathrm{L}^2(\mathbb{R} \times \mathbb{R}^*, |a|^{-2}dbda)$, with

$$\|F_{(b_0,a_0)}\|_2 = \|F\|_2 \quad .$$

The left translates define the **left regular representation** λ_G of G on $\mathrm{L}^2(\mathbb{R}\times\mathbb{R}^*, |a|^{-2}dbda)$, by

$$\lambda_G(b_0, a_0)(F) = F_{(b_0,a_0)} \quad .$$

(c) The CWT $\mathrm{L}^2(\mathbb{R}) \to \mathrm{L}^2(\mathbb{R}\times\mathbb{R}^*, |a|^{-2}dbda)$ is an **intertwining operator** between π and λ_G. This means that

$$W_\psi(\pi(b', a')f) = \lambda_G(b', a')(W_\psi f) \quad .$$

In concrete formulas, this means

$$W_\psi(T_{b'} D_{a'} f) = W_\psi f(b - a'^{-1}b', aa'^{-1}) \quad .$$

(d) The admissibility condition (2.1) also has a representation-theoretic interpretation.

□

2.2 Wavelets and Derivatives I

In the following two sections we want to discuss the connection between decay of wavelet coefficients and (local) smoothness of the analysed function. In order to get an idea of the results to be established, observe that by 1.6.2, we have an almost equivalence

$$f \in C^n(\mathbb{R}) \rightleftharpoons |\widehat{f}(\omega)\omega^n| \to 0 \quad , \quad \text{as } \omega \to 0$$

(see 1.6.2 for the precise formulation). On the other hand, we established in (2.8) an interpretation of the sort

$W_\psi f(b, a) \approx$ Fourier coefficient corresponding to frequency proportional to $1/a$ and position b .

Combining the two heuristics (with proper normalisation), we expect a connection

$$f \in C^n(\text{ near x }) \rightleftharpoons |a^{n+1/2} W_\psi f(b, a)| \to 0 \ , \ \text{for } a \to 0 \text{ and } b \text{ near } x \quad .$$

Even though our argument is only heuristic, there is a way of making it precise. As a key to the proof we require additional properties of the wavelets. The first one is the number of vanishing moments:

Definition 2.2.1 $g : \mathbb{R} \to \mathbb{C}$ **has** n **vanishing moments** ($n \in \mathbb{N}$) if for all $0 \le j < n$:

$$\int_{\mathbb{R}} t^i g(t) dt = 0 \quad ,$$

with absolute convergence. □

Vanishing moments have a simple interpretation on the Fourier transform side: If g has n vanishing moments, then by 1.6.5 $\widehat{g} \in C^n(\mathbb{R})$, and $\widehat{g}^{(i)}(0) = 0$ for all $0 \le i < n-1$. Thus the number of vanishing moments is the order to which $\widehat{g}$ vanishes at 0.

The number of vanishing moments will be a key property in dealing with smooth functions. The following simple argument shows why: Given any ψ with n vanishing moments and any polynomial p of degree less than n, we may write (with some abuse of notation) that $\langle \psi, p \rangle = 0$. Hence, if f is n-fold differentiable at a point x, we can write

$$f = f(x_0) + p(x - x_0) + r(x - x_0) \quad ,$$

where p is the Taylor polynomial of f of degree $n-1$ and $r(x-x_0)$ is a term of order $(x-x_0)^n$. Hence, with a wavelet, say, of compact support nd n vanishing moments, we expect

$$\langle f, T_b D_a \psi \rangle = \langle p + r, T_b D_a \psi \rangle = \langle r, T_b D_a \psi \rangle = O(|a|^{n+1/2})$$

Next we clarify some of the notions which come up when dealing with local smoothness of functions.

Definition 2.2.2 The space $\mathrm{L}^1_{loc}(\mathbb{R})$ consists of all measurable functions $g : \mathbb{R} \to \mathbb{C}$ with the properties, that $g \cdot \chi_K \in \mathrm{L}^1(\mathbb{R})$, for all compact sets $K \subset \mathbb{R}$. It is immediate from the definition that $L^p(\mathbb{R}) \subset L^1_{loc}(\mathbb{R})$ for all $p \ge 1$. Given $f \in \mathrm{L}^1_{loc}(\mathbb{R})$, a function F is called **antiderivative** of f if

$$\forall x, y \in \mathbb{R} : F(x) - F(y) = \int_x^y f(t) dt \quad .$$

Antiderivatives are continuous and unique up to an additive constant. If f has a vanishing moment, $F(x) = \int_{-\infty}^x f(x) dx$ is called the **canonical antiderivative**. If F is differentiable at $x \in \mathbb{R}$, then x is called **Lebesgue point** of f (not of F!). □

The definition of a Lebesgue point is not identical with that in [13]; however, the difference only concerns a set of Lebesgue measure zero. In particular, the next theorem is not affected (see [13, 7.7].)

Theorem 2.2.3 *(Lebesgue Differentiation Theorem) Let* $f \in \mathrm{L}^1_{loc}(\mathbb{R})$*, and define* $M = \{x \in \mathbb{R} : x \text{ is not a Lebesgue point of } f\}$*. Then* M *has measure zero.*

Lemma 2.2.4 *(Partial integration) Let* $f, g \in \mathrm{L}^1_{loc}(\mathbb{R})$ *and* F, G *be antiderivatives. Then, for all* $x, y \in \mathbb{R}$*,*

$$\int_x^y F(t) g(t) dt = (FG)|_x^y - \int_x^y f(t) G(t) dt \quad .$$

Proof. W.l.o.g. $F(x) = G(x) = 0$. Hence we need to show

$$\begin{aligned} 0 &= \int_x^y \int_x^t f(s)g(t)dsdt + \int_x^y \int_x^t f(t)g(s)dsdt - \int_x^y \int_x^y f(t)g(s)dsdt \\ &= \int_x^y \int_x^t f(s)g(t)dsdt - \int_x^y \int_t^y f(t)g(s)dsdt \\ &= \int_{\Delta_1} h_1(s,t)dsdt - \int_{\Delta_2} h_2(s,t)dsdt \quad , \end{aligned}$$

where $\Delta_1 = \{(s,t) : s \le t\}$, $\Delta_2 = \{(s,t) : s \ge t\}$, $h_1(s,t) = f(s)g(t)$ and $h_2(s,t) = f(t)g(s)$. Now a change of variables $\phi : (s,t) \mapsto (t,s)$ shows that the two integrals indeed cancel. □

2.2.5 Exercises (a) Give a measurable function F which is a.e. differentiable, such that the equation

$$F(x) - F(y) = \int_x^y F'(t)dt$$

does not hold.
(b) Give F, G measurable and a.e. differentiable such that partial integration does not hold for F, G, F', G'. □

Definition 2.2.6 Let $f : \mathbb{R} \to \mathbb{C}$ and $\alpha \ge 0$ be given. Then f **has decay rate** α if there exists $C > 0$ such that for all $x \in \mathbb{R}$ the inequality $|f(x)| \le C(1 + |x|)^{-\alpha}$ holds. □

We observe that $f \in \mathrm{L}^p(\mathbb{R})$, for all $p \ge 1$, whenever f has decay rate $\alpha > 1/p$. If $f \in S(\mathbb{R})$, then $f^{(n)}$ has decay rate α, for all $n \ge 0$ and $\alpha > 1$.

Proposition 2.2.7 *(a) Let f be measurable with decay rate $\alpha > n+1$, and with n vanishing moments. Then the canonical antiderivative F of f has $n-1$ vanishing moments and decay rate $\alpha - 1$. We define inductively : F =:* **first canonical antiderivative**, *first canonical antiderivative of F =:* **second canonical antiderivative** *etc.*
(b) Conversely assume that $F \in C^{n-1}(\mathbb{R})$ such that F^{n-1} is almost everywhere differentiable, and F^{n-1} is an antiderivative of $F^{(n)}$. If in addition

$$\int_{\mathbb{R}} |t|^n |F^{(n)}(t)| dt < \infty$$

, then F has n vanishing moments.

Proof. (a) Let us first prove the decay rate. Given $x > 0$,

$$\begin{aligned} |F(x)| &\le \int_x^\infty C(1+|t|)^{-\alpha} dt = \frac{C}{\alpha - 1} t^{1-\alpha}|_{x+1}^{\infty} \\ &= \frac{C(1+x)^{1-\alpha}}{\alpha - 1} \end{aligned}$$

and an analogous argument works for $x < 0$. But this implies that

$$\int_{\mathbb{R}} |t|^n |F(t)| dt < \infty \quad ,$$

as well as

$$\begin{aligned}\int_{\mathbb{R}} t^{n-1}F(t)dt &= \lim_{T\to\infty}\int_{-T}^{T} t^{n-1}F(t)dt \\ &= \lim_{T\to\infty}\left(\frac{t^n F(t)}{n}\Big|_{-T}^{T} - \int_{-T}^{T} t^n f(t)dt\right)\end{aligned}$$

and the first term vanishes because F has decay rate $\alpha - 1 > n$, and the second term vanishes because f has n vanishing moments. Part (b) is proved with n-fold partial integration. □ Note that the assumptions for the proposition are complicated but rather mild. One application of part (b) lies in manufacturing functions with vanishing moments: Simply picking $F \in C_c^n(\mathbb{R})$ and letting $\psi = F^{(n)}$ yields a ψ with n vanishing moments and compact support.

The following lemma is probably not in the most general form. Its function is to clarify when differentiation can be pulled into the convolution product.

Lemma 2.2.8 *Let f, F, g, G be functions with the following properties:*

(i) g has decay rate $\alpha > 1$, $\int_{\mathbb{R}} g(x)dx = 0$, G is the canonical antiderivative of g.

(ii) $f \in \mathrm{L}^2(\mathbb{R})$, and F is an antiderivative of F.

(iii) F is bounded or $\mathrm{supp}(g)$ is compact.

Then $f * g = (F * g)' = (f * G)'$

Proof. We compute for $h > 0$ the difference quotient

$$\begin{aligned}\left|\frac{(f*G)(x) - (f*G)(x-h)}{h} - f*g(x)\right| &= \left|\int_{\mathbb{R}} f(y)\left(\frac{G(x-y) - G(x-h-y)}{h} - g(x-y)\right)dy\right| \\ &\le \|f\|_2 \left(\int_{\mathbb{R}}\left|\frac{G(x-y) - G(x-h-y)}{h} - g(x-y)\right|^2 dy\right)^{1/2} .\end{aligned}$$

The integrand goes pointwise to zero on the Lebesgue set of g, i.e., almost everywhere by 2.2.3. Moreover, assuming w.l.o.g. that $|h| < 1/2$,

$$\begin{aligned}\left|\frac{G(x-y) - G(x-h-y)}{h} - g(x-y)\right| &= \left|\int_{x-y-h}^{x-y}\frac{g(t) - g(x-y)}{h}dt\right| \\ &\le \frac{1}{h}\int_{x-y-h}^{x-y}\sup\{|g(\xi) - g(x-y)| : \xi \in [x-y-h, x-y]\} \\ &\le |g(x-y)| + \sup\{|g(\xi)| : \xi \in [x-y-h, x-y]\} \\ &\le C((1+|x-y|)^{-\alpha} + (1+|x-y|-|h|)^{-\alpha}) \\ &\le C((1+|x-y|)^{-\alpha} + (1/2+|x-y|)^{-\alpha})\end{aligned}$$

shows that the integrand can be uniformly estimated by an integrable function. Thus Lebesgue's dominated convergence theorem applies to yield $(f*G)'(x) = (f*g)(x)$.

Moreover,

$$\begin{aligned}(F*g)(x) &= \lim_{T\to\infty}\int_{-T}^{T} F(y)g(x-y)dy \\ &= \lim_{T\to\infty}\left(-F(T)G(x-T) + F(-T)G(x+T) + \int_f (y)G(x-y)dy\right) .\end{aligned}$$

Now the first two terms vanish by assumption (iii), thus $f * G$ remains. $\square$

Lemma 2.2.9 *(Approximate Identities) Let g be bounded and measurable, f with decay rate $\alpha > 1$. Then*

$$\lim_{a \to 0} (g * a^{-1/2}(D_a f))(x) = g(x) \int_{\mathbb{R}} f(t) dt$$

pointwise whenever g is continuous at x, as well as uniformly on all intervals on which g is uniformly continuous.

Proof. Given $\epsilon > 0$, pick $\delta > 0$ such that $|g(x) - g(x-t)| < \epsilon$ for all $|t| < \delta$. Then

$$\begin{aligned}
\left| g(x) \int_{\mathbb{R}} f(t)dt - g * |a|^{-1/2}(D_a f)(x) \right| &= \left| \int_{\mathbb{R}} g(x)|a|^{-1} f(t/a) dt - \int_{\mathbb{R}} g(x-t)|a|^{-1} f(t/a) dt \right| \\
&= \left| \int_{\mathbb{R}} (g(x) - g(x-t))|a|^{-1} f(t/a) dt \right| \\
&\leq \int_{|t|<\delta} |g(x) - g(x-t)||a|^{-1}|f(t/a)| dt \\
&+ \int_{|t|>\delta} 2\|g\|_\infty C(1 + |t/a|)^{-\alpha} \frac{da}{|a|} \\
&\leq \epsilon \|f\|_1 + 2\|g\|_\infty \int_{|t|>\delta/|a|} C(1+|t|)^{-\alpha} da \quad .
\end{aligned}$$

For fixed a, the first term can be made arbitrarily small. Moreover, since $t \mapsto (1+|t|)^{-\alpha}$ is integrable, the second term tends to zero as $a \to 0$. $\square$

We can now state the first characterisation of smooth functions by use of wavelets. Note that this theorem is a global statement holding for the same set of functions as the Fourier estimates. Thus the full advantage of using wavelet instead of Fourier transform (i.e., compactly supported functions instead of complex exponentials) is not yet visible.

Theorem 2.2.10 *Let $f \in \mathrm{L}^2(\mathbb{R}) \cap C^n(\mathbb{R})$, for some $n \geq 1$. Let ψ be a wavelet with n vanishing moments and decay rate $\alpha > n+1$. Assume in addition that either ($f^{(i)}$ is bounded, for $0 \leq i \leq n-1$) or ($\mathrm{supp}(\psi)$ is compact). Then*

$$a^{-n}|a|^{-1/2} W_\psi f(b,a) \to f^{(n)}(b) \int_{\mathbb{R}} \rho(x) dx$$

where ρ is the nth canonical antiderivative of ψ^*. The convergence is uniform in b on all interval I on which $f^{(n)}$ is uniformly continuous and bounded. In particular on these intervals the wavelet coefficients obey the uniform decay estimate

$$|W_\psi f(b,a)| \leq C_I |a|^{1/2+n} \quad .$$

Proof. By 2.1.11 we have

$$(W_\psi f)(b,a) = \left(f * D_a(\psi^*)\right)(b) = (f * D_a(\rho^{(n)}))(b) \quad .$$

An n-fold application of the chain rule yields $D_a(\rho^{(n)}) = a^n (D_a \rho)^{(n)}$. Thus, applying first 2.2.8 and then 2.2.9, we obtain

$$\begin{aligned}
a^{-n}|a|^{-1/2} W_\psi f(b,a) &= |a|^{-1/2} a^{-n} (f * a^n (D_a \rho)^n)(b) \\
&= |a|^{-1/2} (f^{(n)} * (D_a \rho))(b) \\
&\to f^{(n)}(b) \int_{\mathbb{R}} \rho(x) dx
\end{aligned}$$

This implies both pointwise and uniform convergence. In particular, there exists $a_0 > 0$ such that

$$|W_\psi f(b,a)| \leq |a|^{1/2+n}|1 + f^{(n)}(b)\int_{\mathbb{R}} \rho(x)dx| \leq |a|^{1/2+n}|1 + \|f\|_\infty \int_{\mathbb{R}} \rho(x)dx| \quad ,$$

for all $|a| < a_0$. However, for $|a| > a_0$,

$$|(W_\psi f)(b,a)| \leq \|f\|_2\|g\|_2 \leq \frac{\|f\|_2\|g\|_2}{|a_0|^{n+1/2}}|a|^{n+1/2}$$

hence the uniform estimate is shown. □

2.2.11 Further interpretation

$$\begin{aligned}(W_\psi f)(b,a) &= a^n|a|^{-1/2}(f^{(n)} * D_a\rho) \\ &= |a|^{-1/2}(f * D_a\rho)^{(n)}(b)\end{aligned}$$

yields two new interpretations of CWT. The first sees as $W_\psi f(\cdot, a)$ as smoothed nth derivative of f, with a as smoothing parameter. The second interpretation changes order of differentiation and smoothing, $f \mapsto f * D_a\rho \mapsto (f * D_a\rho)^{(n)}$, which is also applicable to nonsmooth functions f. This will be the next thing on our agenda. □

We next want to discuss local properties of functions in terms of their wavelet coefficients. Note that here finally wavelets will turn out to be superior to the Fourier transform, where an isolated singularity ruins the general picture. Since wavelets can be chosen as compactly supported, the wavelet transform only "sees" portions of the function f of size proportional to a, and thus only the behaviour of f on these portions matters. The following lemma contains a localisation statement which is also valid for rapidly decreasing wavelets instead of compact supported ones, such as the Mexican Hat wavelet.

Lemma 2.2.12 *Let $\psi \in \mathrm{L}^2(\mathbb{R})$ be a wavelet, $f \in \mathrm{L}^2(\mathbb{R})$ with $f|_I = 0$ for a compact interval $[x_0, x_1]$.*

(a) Falls supp$(f) \subset [-T, T]$*, define $M = \{(b,a) : x_0 + |a|T \leq b \leq x_1 - |a|T\}$. Then $W_\psi f(b,a) = 0$, for all $(b,a) \in M$.*

(b) Suppose that ψ has decay rate $\alpha > 1/2$. For $T > 0$ define

$$M = \{(b,a) \in I \times \mathbb{R}^* : x_0 + \sqrt{|a|T} \leq b \leq x_1 - \sqrt{|a|T}\} \quad .$$

Then there exists a constant C, depending on f and T, such that $|W_\psi f(b,a)| \leq C|a|^{\alpha/2-1/4}$, for all $(b,a) \in M$.

Proof. In the case (a), we observe that $(b,a) \in M$ implies supp$(T_b D_a \psi) \subset I$, and thus $W_\psi f(b,a) = 0$ by assumption. In the second case, we first compute

$$\begin{aligned}|W_\psi f(b,a)| &= \left| \int_{\mathbb{R}\setminus I} f(x)|a|^{-1/2}\overline{\psi\left(\frac{x-b}{a}\right)}dx \right| \\ &\leq \|f\|_2 \|(T_b D_a\psi) \cdot (1 - \chi_I)\|_2\end{aligned}$$

by the Cauchy-Schwarz inequality. Define $s = min(b - x_0, x_1 - b)$, for $(b,a) \in M$ this implies $s \geq \sqrt{|a|T}$. Moreover,

$$\int_{\mathbb{R}} |a|^{-1} \left|\psi\left(\frac{x-b}{a}\right)\right|^2 dx = \int_{\mathbb{R}\setminus[(x_0-b)/a,(x_1-b)/a]} |\psi(x)|^2 dx \leq \int_{|x|\geq s/|a|} |\psi(x)|^2 dx$$

since by choice of s we have $[-s/|a|, s/|a|] \subset [(x_0-b)/a, (x_1-b)/a]$. Using the decay rate of ψ, we can estimate

$$\int_{|x|\geq s/|a|} |\psi(x)|^2 dx \leq C(s/|a|)^{1-2\alpha} \leq C(\sqrt{|a|T}/|a|)^{1-2\alpha} \leq C'|a|^{\alpha-1/2}$$

Therefore $\|(T_b D_a \psi)\cdot(1-\chi_I)\|_2 \leq C|a|^{\alpha/2-1/4}$. □

We can now apply the localisation argument to extend the theorem to more general functions.

Theorem 2.2.13 *Let $f \in \mathrm{L}^2(\mathbb{R})$, $I = [x_0, x_1] \subset \mathbb{R}$ a compact interval. Let f be a function that is in (x_0, x_1) n times continuously differentiable, and such that*

$$\lim_{x\to x_0, x>x_0} f^{(n)}(x) \quad , \quad \lim_{x\to x_1, x<x_1} f^{(n)}(x)$$

exist. Let ψ be a wavelet with n vanishing moments, and assume that either suppψ *is compact or that ψ has decay rate $\alpha > 2n + 3/2$. Then*

$$a^{-n}|a|^{-1/2} W_\psi f(b,a) \to f^{(n)}(b) \int_{\mathbb{R}} \rho(x)dx \quad , \quad \textit{as } a \to 0$$

and

$$|W_\psi f(b,a)| \leq C|a|^{1/2+n}$$

both hold uniformly in M, where M is the set defined in the previous lemma.

Proof. The existence of $\lim_{x\to x_0,x>x_0} f^{(n)}(x)$ and $\lim_{x\to x_1,x<x_1} f^{(n)}(x)$ imply that $f^{(n)}$ is bounded and uniformly continuous. Moreover, it is easily shown (inductively) that $\lim_{x\to x_0,x>x_0} f^{(i)}(x)$ and $\lim_{x\to x_1,x<x_1} f^{(i)}(x)$ also exist, for $0 \leq i < n$. Then $f|_I$ can be extended to a C_c^n-function $\widetilde{f}$. By 2.2.10 the statements of the theorem hold for $\widetilde{f}$. By Lemma 2.2.12 the error in replacing f by $\widetilde{f}$ can be estimated by

$$|a|^{-n-1/2}|W_\psi f(b,a) - W_\psi \widetilde{f}(b,a)| \leq C|a|^{\alpha-1/2}|a|^{-n-1/2} = C|a|^{\alpha-n-1} \to 0 \quad .$$

This concludes the proof. □

2.3 Wavelets and Derivatives II

In this section we want to show a converse of Theorem 2.2.13, i.e., wavelet coefficient decay entails smoothness. As could be expected, pointwise wavelet inversion will play a role in this context. First we need a statement about approximate units which is somewhat more general than what we showed in 2.2.9.

Theorem 2.3.1 *Let $f \in \mathrm{L}^2(\mathbb{R})$, $g : \mathbb{R} \to \mathbb{C}$ with decay order $\alpha > 1$. Then*

$$\lim_{a\to 0} \left[f * (D_a g)\right](x) = f(x) \int_{\mathbb{R}} g(t)dt$$

at all Lebesgue points of $\mathbb{R}$.

Proof. For the case that $\int_{\mathbb{R}} g(t)dt \neq 0$ see [15, 9.13]. The general case is easily concluded from this. □

Theorem 2.3.2 *(Pointwise wavelet inversion)*
Let $\psi \in \mathrm{L}^2(\mathbb{R})$ be a bounded wavelet with compact support. For $f \in \mathrm{L}^2(\mathbb{R})$ and $\epsilon > 0$ define

$$f_\epsilon(x) = \frac{1}{c_\psi} \int_{|a|>\epsilon} \int_{\mathbb{R}} (W_\psi f)(b,a)\ T_b D_a \psi(x)\ db \frac{da}{|a|^2} \quad .$$

Here the right-hand side converges absolutely. Then, for all Lebesgue-points x of f, $\lim_{\epsilon\to 0} f_\epsilon(x) = f(x)$.

Proof. Step 1. Absolute convergence of the integral

$$\begin{aligned} \int_{|a|>\epsilon} \int_{\mathbb{R}} |W_\psi f(b,a)| |T_b D_a \psi(x)| db \frac{da}{|a|^2} &\leq \|f\|_2 \|\psi\|_2 \int_{|a|>\epsilon} \int_{\mathbb{R}} \left|\psi\left(\frac{x-b}{a}\right)\right| db \frac{da}{|a|^{5/2}} \\ &= \|f\|_2 \|\psi\|_2 \int_{|a|>\epsilon} \|\psi\|_1 \frac{da}{|a|^{3/2}} < \infty \quad . \end{aligned}$$

Step 2. $f_\epsilon = f * G_\epsilon$, where

$$G_\epsilon(t) = \frac{1}{c_\psi} \int_{|a|<1/\epsilon} (\psi^* * \psi)(at) da \quad .$$

Indeed,

$$\begin{aligned} f_\epsilon(x) &= \frac{1}{c_\psi} \int_{|a|>\epsilon} \int_{\mathbb{R}} \int_{\mathbb{R}} f(y) \overline{|a|^{-1/2} \psi\left(\frac{y-b}{a}\right)} dy |a|^{-1/2} \psi\left(\frac{x-b}{a}\right) db \frac{da}{|a|^2} \\ &= \frac{1}{c_\psi} \int_{\mathbb{R}} f(y) \int_{|a|>\epsilon} \int_{\mathbb{R}} |a|^{-1} \psi^*\left(\frac{b-y}{a}\right) \psi\left(\frac{x-b}{a}\right) db \frac{da}{|a|^2} dy \\ &= \frac{1}{c_\psi} \int_{\mathbb{R}} f(y) v \int_{|a|>\epsilon} \int_{\mathbb{R}} \psi^*(b) \psi\left(\frac{x-y}{a} - b\right) db \frac{da}{|a|^2} dy \\ &= (f * \widetilde{G}_\epsilon)(x) \quad , \end{aligned}$$

where

$$\begin{aligned} \widetilde{G}_\epsilon(t) &= \frac{1}{c_\psi} \int_{|a|>\epsilon} \int_{\mathbb{R}} \psi^*(b) \psi\left(\frac{t}{a} - b\right) db \frac{da}{|a|^2} \\ &= \frac{1}{c_\psi} \int_{|a|>\epsilon} (\psi^* * \psi)\left(\frac{t}{a}\right) \frac{da}{|a|^2} \\ &= \frac{1}{c_\psi} \int_{|a|<1/\epsilon} (\psi^* * \psi)(at) da \\ &= G_\epsilon(t) \quad . \end{aligned}$$

Step 3. $G_\epsilon = \epsilon^{-1/2}(D_\epsilon G_1)$.

$$\begin{aligned} G_\epsilon(t) &= \frac{1}{c_\psi}\int_{|a|<1/\epsilon}(\psi^* * \psi)(at)da \\ &= \frac{1}{c_\psi}\cdot\frac{1}{\epsilon}\int_{|a|<1}(\psi^* * \psi)\left(\frac{at}{\epsilon}\right)da \\ &= 1/\epsilon G_1(t/\epsilon) \quad . \end{aligned}$$

Step 4. G_1 is continuous with compact support.
Indeed, suppose that supp$\psi \subset [-T,T]$, then supp$(\psi^* * \psi) \subset [-2T, 2T]$. Thus, for all t with $|t| > 2T$,

$$\begin{aligned} G_1(t) &= \frac{1}{c_\psi}\int_{|a|<1}(\psi^* * \psi)(at)da \\ &= \frac{1}{c_\psi}\frac{1}{|t|}\int_{|a|<t}(\psi^* * \psi)(a)da \\ &= \frac{1}{c_\psi}\frac{1}{|t|}\int_{\mathbb{R}}(\psi^* * \psi)(a)da \\ &= \frac{1}{|t|c_\psi}\mathcal{F}(\psi^* * \psi)(0) \\ &= \frac{1}{|t|c_\psi}|\widehat{\psi}(0)|^2 \\ &= 0 \quad . \end{aligned}$$

Moreover, $\psi^* * \psi$ is uniformly continuous, and thus G_1 is continuous.
Step 5. $\int_{\mathbb{R}} G_1(t)dt = 1$.
For this purpose we show that

$$\widehat{G_1}(t) = \frac{1}{c_\psi}\int_t^\infty \frac{|\widehat{\psi}(\omega)|^2 + |\widehat{\psi}(-\omega)|^2}{\omega}d\omega \quad . \tag{2.10}$$

Note that once this equality is established, we obtain $\widehat{G_1}(0) = 1$ by taking the limit $t \to 0$ (observing that $G_1 \in \mathrm{L}^1(\mathbb{R})$, thus $\widehat{G_1}$ is continuous), and Step 5 is proved.

For the proof of (2.10 observe that G_1 solves the differential equation

$$\frac{d}{dt}[t \cdot G_1(t)] = \frac{2}{c_\psi}\mathrm{Re}(\psi^* * \psi)(t) \quad .$$

Indeed, for $t > 0$,

$$\begin{aligned} t \cdot G_1(t) &= \frac{1}{c_\psi}\int_0^t (\psi^* * \psi)(x) + (\psi^* * \psi)(-x)dx \\ &= \frac{2}{c_\psi}\int_0^t 2\mathrm{Re}(\psi^* * \psi)(x)dx \quad , \end{aligned}$$

since $(\psi^* * \psi)^* = (\psi^* * \psi)$. This proves the differential equation for positive t. For negative t it follows since G_1 is an even function, and hence $[t \cdot G_1(t)]'$ is even, as is $\mathrm{Re}(\psi^* * \psi)$.

Now, on the Fourier transform side, using 1.6.1, 1.6.5 and the relation

$$F(\overline{g})(\omega) = \overline{\widehat{g}(-\omega)}$$

the differential equation translates to

$$-\omega \widehat{G_1}'(\omega) = \frac{1}{c_\psi} \left(|\widehat{\psi}(\omega)|^2 + |\widehat{\psi}(-\omega)|^2 \right)$$

hence solving for $\widehat{G_1}$ gives

$$\widehat{G_1}(t) = \frac{1}{c_\psi} \int_t^\infty \frac{|\widehat{\psi}(\omega)|^2 + |\widehat{\psi}(-\omega)|^2}{\omega} d\omega + C \quad .$$

$G_1 \in \mathrm{L}^1(\mathbb{R})$ implies $\widehat{G_1}(t) \to 0$ as $t \to \infty$, and thus $C = 0$. This proves 2.10.
Step 6. Summarising: $f_\epsilon = f * \epsilon^{-1/2}(D_\epsilon G_1)$, where G_1 is bounded with compact support and $\int_{\mathbb{R}} G_1(x)dx = 1$. Thus Theorem 2.3.1 implies the convergence statement. □

It is remarkable to note that the most time-consuming step in this proof was to show that G_1 has the correct normalisation.

Now we can finally prove the converse of Theorem 2.2.13. Note how the assumptions on the wavelet have changed: No vanishing moment condition is needed; instead, we require smoothness.

Theorem 2.3.3 *Let $f, \psi \in \mathrm{L}^2(\mathbb{R})$, where ψ is admissible with* $\mathrm{supp}(\psi) \subset [-T, T]$, $\psi \in C^{n+2}$. *Given a compact interval $[x_0, x_1] \subset \mathbb{R}$ and $a_0 > 0$,define the set*

$$N_T = \{(b, a) : x_0 - |a|T < b < x_1 + |a|T, |a| < a_0\} \quad .$$

Assume that the exists a constance $C > 0$ such that for all $(b, a) \in N_T$, $|W_\psi f(b, a)| \le C|a|^\alpha$, with $\alpha > n + 1/2$. Then f is in (x_0, x_1) n times continuously differentiable.

Proof. The choice of the set N_T is motivated by the following calculation. Obviously, for the reconstruction of $f(x)$, only those (b, a) contribute for which $x \in \mathrm{supp}(T_b D_a \psi)$. Thus, if we let $0 < \epsilon < \delta < a_0$ and define f_ϵ and f_δ according to Theorem 2.3.2, we can compute

$$\begin{aligned} f_\epsilon(x) - f_\delta(x) &= \frac{1}{c_\psi} \int_{\epsilon<|a|<\delta} \int_{\mathbb{R}} (W_\psi f)(b, a)(T_b D_a \psi)(x) db \frac{da}{|a|^2} \\ &= \frac{1}{c_\psi} \int_{\epsilon<|a|<\delta} \int_{x-|a|T}^{x+|a|T} (W_\psi f)(b, a)(T_b D_a \psi)(x) db \frac{da}{|a|^2} \end{aligned}$$

and $[x - |a|T, x + |a|T] \times \{a\} \subset N_T$, for all $|a| < a_0, x \in [x_0, x_1]$. Hence the estimate

$$|W_\psi f(b, a)| \le C|a|^\alpha$$

applies to the integrand.

We now proceed the proof of the statement by induction over n.
The case $n = 0$: Plugging in the estimate, we obtain

$$\begin{aligned} |f_\epsilon(x) - f_\delta(x)| &\le \frac{1}{c_\psi} \int_{\epsilon<|a|<\delta} \int_{\mathbb{R}} |a|^\alpha |a|^{-5/2} \left| \psi \left(\frac{x - b}{a} \right) \right| db da \\ &= \frac{1}{c_\psi} \|\psi\|_1 \int_{\epsilon<|a|<\delta} |a|^{\alpha - 3/2} da \quad . \end{aligned}$$

This holds uniformly in $x \in [x_0, x_1]$. The assumption $\alpha > 1/2$ implies that $\int_0^{a_0} |a|^{\alpha-3/2} da < \infty$, hence we obtain the uniform limit

$$\lim_{\delta \to 0, \epsilon < \delta} |f_\epsilon(x) - f_\delta(x)| = 0 \quad .$$

Moreover, as seen in Step 2. of the Proof of 2.3.2, $f_\epsilon = f * G_\epsilon$ is continuous. Thus the family $(f_\epsilon)_{\epsilon<1}$ consists of continuous functions which are Cauchy with respect to the sup-norm. Thus there exists a uniform limit $\widetilde{f}$ to which the f_ϵ converge; which is necessarily continuous. On the other hand $f_\epsilon \to f$ on all Lebesgue points, therefore $f = \widetilde{f}$ is continuous.

Induction step: We first need a few auxiliary statements concerning the derivatives of f_ϵ.

Part (a). For $0 \le l \le n$,

$$f_\epsilon^{(l)}(x) = \frac{1}{c_\psi} \int_{|a|<\epsilon} \int_{\mathbb{R}} W_\psi f(b,a) a^{-l} |a|^{-1/2} \psi^{(l)} db \frac{da}{|a|^2}$$

We prove this by induction. The case $l = 0$ is the definition of f_ϵ. Given the induction hypothesis for $l \ge 0$, define

$$g_\epsilon(x) = \frac{1}{c_\psi} \int_{|a|>\epsilon} \int_{\mathbb{R}} W_\psi f(b,a) a^{-(l+1)} |a|^{-1/2} \psi^{(l+1)} \left(\frac{x-b}{a} \right) db \frac{da}{|a|^2}$$

Using the induction hypothesis we have

$$\frac{f_\epsilon^{(l)}(x) - f_\epsilon^{(l)}(x-h)}{h} - g_\epsilon(x) = \frac{1}{c_\psi} \int_{|a|>\epsilon} (W_\psi f)(b,a) a^{-l} |a|^{-1/2} \left(\frac{\psi^{(l)}\left(\frac{x-b}{a}\right) \psi^{(l)}\left(\frac{x-b-h}{a}\right)}{h} - a^{-1} \psi^{(l+1)} \left(\frac{x-b}{a} \right) \right) db \frac{da}{|a|^2}$$

where the integrand goes to zero pointwise, as h goes to zero. Moreover, assuming $|h| < 1$ we can estimate the integrand as

$$\begin{aligned} & \left| (W_\psi f)(b,a) a^{-l} |a|^{-1/2} \left(\frac{\psi^{(l)}\left(\frac{x-b}{a}\right) \psi^{(l)}\left(\frac{x-b-h}{a}\right)}{h} - a^{-1} \psi^{(l+1)} \left(\frac{x-b}{a} \right) \right) \right| \\ & \le \|f\|_2 \|\psi\|_2 |a|^{-l-5/2} \sup\{ |\psi^{(l+2)}(t/a)| : t \text{ between } x-b-1, x-b+1 \} \end{aligned}$$

Now it is not hard to see that the right hand side is an integrable function, thus the dominated convergence theorem applies to yield the desired statement.

Part (b). $(f_\epsilon^{(n+1)})_{\epsilon<1}$ is uniformly Cauchy in $[x_0, x_1]$.

For $\epsilon < \delta < a_0$ we have

$$\begin{aligned} |f_\epsilon^{(n+1)}(x) - f_\delta^{(n+1)}(x)| & \le \frac{1}{c_\psi} \int_{\epsilon<|a|<\delta} \int_{\mathbb{R}} |a|^\alpha |a|^{-n-1} |a|^{-1/2} \left| \psi^{(n+1)} \left(\frac{x-b}{a} \right) \right| db \frac{da}{|a|^2} \\ & = \frac{1}{c_\psi} \|\psi^{(n+1)}\|_1 \int_{\epsilon<|a|<\delta} |a|^{\alpha-(n+1)-5/2} da \\ & \to 0 \quad , \end{aligned}$$

as $\delta \to 0$. To see the latter convergence, note that by assumption $\alpha > (n+1) + 3/2$, and thus $\alpha - (n+1) - 5/2 > -1$. But then

$$\int_0^{a_0} |a|^{\alpha-(n+1)-5/2} da < \infty \quad .$$

Part (c). $(f_\epsilon^{(n+1)})_{\epsilon<1}$ consists of continuous functions.
This follows by similar arguments as in the proof of Theorem 2.3.2: $f_\epsilon^{(n+1)} = f * H_\epsilon$, with

$$H_\epsilon(x) = \frac{1}{c_\psi} \int_{|a|>\epsilon} a^{-(n+1)} (\psi^* * \psi^{(n+1)})(at) da \quad .$$

As in Step 2. of that proof one can show that H_ϵ is continuous with compact support. Hence $f_\epsilon^{(n+1)}$ is continuous.
Part (d). By Parts (b) and (c), $g = \lim_{\epsilon\to 0} f_\epsilon^{(n+1)}$ exists with uniform convergence in $[x_0, x_1]$, and thus g is continuous. Hence finally

$$\begin{aligned}
&\left| \frac{f^{(n)}(x) - f^{(n)}(x-h)}{h} - g(x) \right| \\
&\leq \left| \frac{f_\epsilon^{(n)}(x) - f_\epsilon^{(n)}(x-h)}{h} - g(x) \right| + \left| \frac{f^{(n)}(x) - f_\epsilon^{(n)}(x)}{h} \right| + \left| \frac{f^{(n)}(x-h) - f_\epsilon^{(n)}(x-h)}{h} \right| \\
&\leq \sup\{|f_\epsilon^{(n+1)}(t) - g(x)||t-x| < h\} + \frac{2}{h}\|f^{(n)} - f_\epsilon^{(n)}\| \\
&\leq \sup\{|f_\epsilon^{(n+1)}(t) - g(t)||t-x| < h\} + \frac{2}{h}\|f^{(n)} - f_\epsilon^{(n)}\| \frac{2}{h}\|f^{(n)} - f_\epsilon^{(n)}\| + \\
&\quad \sup\{|g(t) - g(x)||t-x| < h\} \quad .
\end{aligned}$$

Given fixed h, this holds for all ϵ, thus we finally obtain that

$$\left| \frac{f^{(n)}(x) - f^{(n)}(x-h)}{h} - g(x) \right| \leq \sup\{|g(t) - g(x)| : |t-x| < h\} \quad .$$

But now continuity of g implies that the right hand side goes to zero as $h \to 0$, thus we have shown that $f^{(n+1)} = g$. □

2.4 Singularities

In this section we want to sketch how wavelet analysis can be employed to look at points where a function is less regular (by comparison to neighbouring points). The result is not phrased in the form of a theorem; it is rather a heuristic than a rigourous mathematical statement. However, the principle serves as a motivation for many algorithms using wavelets, and it uses both the necessary and sufficient conditions that we have established so far. The main definition is the following:

Definition 2.4.1 Let $f \in \mathrm{L}^2(\mathbb{R})$, $\psi \in \mathrm{L}^2(\mathbb{R})$ admissible. A point $(b, a) \in \mathbb{R} \times \mathbb{R}^*$ is called **maximum position** if b is a (local) maximum of the function $|W_\psi f(\cdot, a)|$. Here we only accept maxima which are strict either in a left or right neighborhood. □

Remark 2.4.2 We consider the following setup: $f : [x_0, x_1] \to \mathbb{C}$, and $z \in (x_0, x_1)$ a point with the following properties:

$$f|_{(x_0,z)} \in C^n, f|_{(z,x_1)} \in C^n, but f|_{(x_0,x_1)} \notin C^{(n-1)}$$

One possible example could be two C^1-curves with a jump at point z. Next we pick a compactly supported wavelet with enough vanishing moments and smoothness, such that Theorems 2.2.13 and 2.3.3 are applicable. Let $T > 0$ be such that $\text{supp}(\psi) \subset [-T, T]$. We define sets

$$N = \{(b, a) : z - |a|T \leq b \leq z + |a|T\}$$

and

$$M = \{(b, a) : x_0 + |a|T \leq b \leq x_1 - |a|T\} \quad .$$

Then N consists of two cones (upper and lower half) with vertex sitting at $(z, 0)$. Fix $k \in \mathbb{N}$. Since f is smooth outside of z, Theorem 2.2.13 implies that in $M \setminus N$

$$|W_\psi f(b, a)| \leq C|a|^{1/2+n} \quad .$$

On the other hand, Theorem 2.3.3 implies that this relation cannot hold on all $M \cap \{(b, a) : |a| \leq 1/k\}$, since otherwise we could conclude $f \in C^{n-1}$. Hence there exists a pair $(\widetilde{b_k}, a_k) \in N$ with $|a_k| < 1/k$, and $W_\psi f(\widetilde{b_k}, a_k) > C|a_k|^{1/2+n}$. Now pick a maximum of the function $|W_\psi f(\cdot, a_k)|$ on the interval $[z - |a_k|T, z + |a_k|T]$, which yields a point (b_k, a_k). Then obviously

$$W_\psi(b_k, a_k) \geq W_\psi(\widetilde{b_k}, a_k) > C|a_k|^{1/2+n} \quad .$$

Summarising, we have obtained a sequence $(b_k, a_k)_{k\in\mathbb{N}}$ of maxima positions with the following properties:

(a) $|W_\psi f(b_k, a_k)| > C|a_k|^{1/2+n}$, for all k.

(b) $a_k \to 0$, $b_k \to z$.

Moreover, if $(b'_k, a'_k)_{k\in\mathbb{N}}$ is another sequence of maxima positions with $a'_k \to 0$, $b'_k \to y \neq z$, then we find for all k large enough that $(a'_k, b'_k) \notin N$, therefore $|W_\psi f(a'_k, b'_k)| \leq C|a'_k|^{1/2+n}$. Hence z is uniquely characterised by (a) and (b).
(This argument can be extended to other smoothness classes, such as Hölder regularity. See [7].) □

2.4.3 Maxima Lines If $\psi \in \mathrm{L}^2(\mathbb{R}) \cap C^2(\mathbb{R})$, all maxima positions of $W_\psi f$ are solutions of the (in-)equalities

$$\frac{d}{db}|W_\psi f(b, a)|^2 = 0, \frac{d^2}{d^2b}|W_\psi f(b, a)|^2 < 0 \quad . \tag{2.11}$$

But then (2.11) can be solved for b as a function of a, i.e., there exists a continuous mapping $\phi : [a_0 - \epsilon, a_0 + \epsilon] \to \mathbb{R}$ such that $\{(\phi(a), a) : a \in [a_0 - \epsilon, a_0 + \epsilon]\}$ consists of maxima positions. In other words, (a_0, b_0) lies on a **maxima line**. If we regard only maxima positions as significant which lie on such maxima lines, we obtain the following "algorithm" for the detection of singularities:

1. Compute the maxima lines.

2. Follow those maxima lines $(\phi(a), a)_a$, that propagate towards the axis $a = 0$.

3. The limits $(z, 0)$ of such maxima lines are candidates for singularities.

4. If $(z, 0)$ is the limit of the maxima line $(\phi(a), a)$, use the decay behaviour of $W_\psi f(\phi(a), a)$ to decide whether z is a singularity.

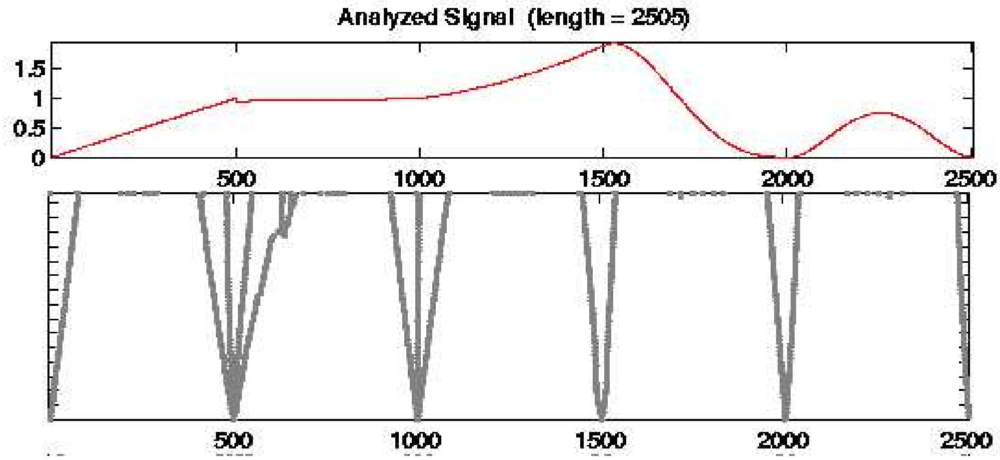

Figure 2.10: Maxima lines for the test signal.

This algorithm is implemented in matlab: Maxima lines are displayed in the CWT gui.

□

2.4.4 Heuristic justification of the maxima line algorithm The robust detection of singularities is an old problem in signal and image analysis. It turns out that the algorithm we have derived above is very similar to a procedure explaining edge detection in the human vision by Marr [8], albeit without any reference to smoothness estimates. An alternative justification may be obtained as follows:

A classical candidate for the detection of sudden jumps is the differentiation operator. Local maxima of the derivative are candidates for jumps. However, often this operation yields a large number of such candidates, and a criterion to discriminate these is needed. A possible way to reduce this number consists in first smoothing the signal, say by some dilated bump ρ. Here the dilation parameter controls the size of ρ and thus the amount of smoothing. Thus we are looking of the local maxima of the function

$$(f * D_a\rho)' \quad .$$

Comparing this with Remark 2.2.11 we find that up to normalisation, this is just a continuous wavelet transform! Now, as the scaling a increases, the smoothing increases, and the jumps detected at scale a are less likely to be at their true positions. Conversely, small scales produce more maxima. Instead of looking for an "optimal" scale a which gives the best tradeoff between these two undesirable effects, the algorithm consists in looking for maxima positions that build chains across scales. This can be seen as an attempt to combine information on several scales to obtain a more robust detector of singularities. □

2.5 Summary: Desirable properties of wavelets

In this chapter we have encountered three properties of wavelets (in varying combination):

1. Smoothness: $\psi \in C^n$

2. Vanishing moments
3. Decay rate

These properties may be summarised under the slogan of "time-frequency-localisation": 1. describes decay of $\widehat{\psi}$ as $\omega \to \infty$, 2. describes such a decay for $\omega \to 0$. 3. obviously expresses decay of ψ for $t \to \infty$.

Chapter 3

Wavelet ONB's and Multiresolution Analysis

A **wavelet ONB** of $L^2(\mathbb{R})$ is by definition an ONB $(\psi_{j,k})_{j\in\mathbb{Z},k\in\mathbb{Z}}$ of $L^2(\mathbb{R})$ of the form

$$\psi_{j,k} = D_{2^j} T_k \psi \quad ,$$

for a suitable choice of $\psi \in L^2(\mathbb{R})$. ψ is often called a **discrete wavelet**. While this definition is fairly simple, it is not clear that such functions ψ exist. We have already encountered one example (the Haar wavelet), but whether there are more such functions is not at all obvious. In this chapter we discuss a construction, called **multiresolution analysis**, which allows to construct such discrete wavelets in a rather systematical fashion; moreover, desirable properties such as smoothness, vanishing moments and compact support may also be guaranteed.

Discrete wavelet systems give rise to discrete wavelet transforms: Every wavelet ONB induces a unitary map

$$W_\psi^d : f \mapsto (\langle f, \psi_{j,k}\rangle)_{j,k\in\mathbb{Z}}$$

with inversion formula

$$f = \sum_{j,k\in\mathbb{Z}} \langle f, \psi_{j,k}\rangle \psi_{j,k} \quad .$$

Thus we have a discrete analog of the isometry property (Theorem 2.1.6) and inversion formula (Theorem 2.1.7) of the continuous wavelet transform. But the analogy goes further: By a result of Daubechies [3] it is known that every ψ giving rise to a wavelet ONB is also admissible in the sense of Definition 2.1.1, and thus

$$W_\psi^d f(j,k) = \langle f, D_{2^j} T_k \psi\rangle = W_\psi f(2^j k, 2^j) \quad .$$

Hence the discrete wavelet transform is nothing but the continuous wavelet transform, sampled on the **dyadic grid**

$$\Gamma = \{(2^j k, 2^j) : j,k \in \mathbb{Z}\} \quad .$$

In particular, we can conclude from the last chapter that discrete wavelet coefficients of piecewise smooth functions vanish quickly, sufficiently many vanishing moments provided.

3.1 Multiresolution analysis

Definition 3.1.1 A **multiresolution analysis (MRA)** is a sequence of closed subspaces $(V_j)_{j\in\mathbb{Z}}$ of $\mathrm{L}^2(\mathbb{R})$ with the following list of properties:

(M1) $\forall j \in \mathbb{Z} : V_j \subset V_{j+1}$.

(M2) $\overline{\bigcup_{j\in\mathbb{Z}} V_j} = \mathrm{L}^2(\mathbb{R})$.

(M3) $\bigcap_{j\in\mathbb{Z}} V_j = \{0\}$.

(M4) $\forall j \in \mathbb{Z}, \forall f \in \mathrm{L}^2(\mathbb{R}) \ : \ (f \in V_j \Leftrightarrow D_{2^j}(f) \in V_0)$.

(M5) $f \in V_0 \Rightarrow \forall m \in \mathbb{Z} : T_m f \in V_0$.

(M6) There exists $\varphi \in V_0$ such that $(T_m\varphi)_{m\in\mathbb{Z}}$ is an ONB of V_0. φ is called **scaling function** of $(V_j)_{j\in\mathbb{Z}}$.

□

The properties (M1) - (M6) are somewhat redundant. They are just listed for a better understanding of the construction. The following discussion will successively strip down (M1)-(M6) to the essential.

Remarks 3.1.2 (a) Properties $(M4)$ and $(M6)$ imply that the scaling function φ uniquely determines the multiresolution analysis: We have

$$V_0 = \overline{\mathrm{span}(T_k\varphi : k \in \mathbb{Z})}$$

by $(M6)$ (which incidentally takes care of $(M5)$) and

$$V_j = D_{2^{-j}} V_0$$

is prescribed by $(M4)$. What is missing are criteria for $(M1) - (M3)$ and $(M6)$. However, note that in the following the focus of attention shifts from the spaces V_j to the scaling function φ.

(b) The parameter j can be interpreted as **resolution** or **(inverse) scale** or (with some freedom) **frequency** parameter. Thus the inclusion property $(M1)$ has a quite natural interpretation: Increasing resolution amounts to adding information.

If we denote by P_j the projection onto V_j, we obtain the characterisations

$$\begin{array}{lcl} (M2) & \Leftrightarrow & \forall f \in \mathrm{L}^2(\mathbb{R}) : \|f - P_j f\| \to 0 \ , \quad \text{as } j \to \infty \\ (M3) & \Leftrightarrow & \forall f \in \mathrm{L}^2(\mathbb{R}) : \|P_j f\| \to 0 \ , \quad \text{as } j \to -\infty \end{array}$$

$P_j f$ can be interpreted as an approximation to f with resolution 2^j. Thus $(M2)$ implies that this approximation converges to f, as resolution increases. □

3.1.3 Exercise ("Shannon-MRA") Let

$$V_0 = \{f \in \mathrm{L}^2(\mathbb{R}) : \mathrm{supp}(\widehat{f}) \subset [-1/2, 1/2]\}$$

and $\varphi(x) = \mathrm{sinc}(x) := \frac{\sin \pi x}{\pi x}$. Prove that

(a) $(T_k\varphi)_{k\in\mathbb{Z}}$ is an ONB of V_0. (Hint: Use the Fourier transform.)

(b) V_0 consists of continuous functions, with

$$\langle f, T_k\varphi\rangle = f(k) \ , \quad \forall k \in \mathbb{Z} \ , \ \forall f \in V_0 \ .$$

Thus we obtain an expansion

$$f(x) = \sum_{k\in\mathbb{Z}} f(k)\mathrm{sinc}(x-k)$$

with convergence both in $\mathrm{L}^2(\mathbb{R})$ and uniformly. (Hint for uniform convergence: Use that $\widehat{f} \in \mathrm{L}^2([-1/2,1/2]) \subset \mathrm{L}^1(\mathbb{R})$.)

(c) Defining $V_j = D_{2^{-j}}V_0$, we obtain

$$V_j = \{f \in \mathrm{L}^2(\mathbb{R}) : \mathrm{supp}(\widehat{f}) \subset [-2^{j-1}, 2^{j-1}]\} \quad .$$

Conclude that $(V_j)_{j\in\mathbb{Z}}$ is an MRA.

□

Using dilation one can show more generally for $f \in \mathrm{L}^2(\mathbb{R})$ with $\mathrm{supp}(\widehat{f}) \subset [-\Omega/2, \Omega/2]$ the **sampling expansion**

$$f(x) = \sum_{k\in\mathbb{Z}} f\left(\frac{k}{\Omega}\right) \mathrm{sinc}\left(\frac{x-k}{\Omega}\right) \quad .$$

Ω is called **bandwidth** f, whereas $1/\Omega$ is called **sampling rate**. The fact that bandwidth and sampling rate are inverse of each other can be seen as one motivation for the use of the dyadic grid in sampling wavelet transforms.

We now start to discuss criteria for the properties $(M1)-(M6)$. As we already remarked, it is enough to do this in terms of the scaling function.

Lemma 3.1.4 *(Characterisation of $(M6)$.)*
Let $\varphi \in \mathrm{L}^2(\mathbb{R})$, and $V_0 = \overline{\mathrm{span}(T_k\varphi : k \in \mathbb{Z})}$. Then

$$(T_k\varphi)_{k\in\mathbb{Z}} \text{ ONB of } V_0 \Leftrightarrow \sum_{\ell\in\mathbb{Z}} |\widehat{\varphi}(\psi+\ell)|^2 = 1 \quad , \ a.e. \ \psi \in \mathbb{R} \tag{3.1}$$

Proof. Using 1.2.8, we know that

$$\begin{aligned}
(T_k\varphi)_{k\in\mathbb{Z}} \text{ ONB of } V_0 \quad &\Leftrightarrow \quad (T_k\varphi)_{k\in\mathbb{Z}} \text{ ONS} \\
&\Leftrightarrow \quad \forall k \in \mathbb{Z} : \langle T_k\varphi, \varphi\rangle = \delta_{k,0} \\
&\Leftrightarrow \quad \forall k \in \mathbb{Z} : \langle \widehat{T_k\varphi}, \widehat{\varphi}\rangle = \delta_{k,0} \\
&\Leftrightarrow \quad \forall k \in \mathbb{Z} : \delta_{k,0} = \int_{\mathbb{R}} e^{-2\pi i\omega k}\widehat{\varphi}(\omega)\overline{\widehat{\varphi}(\omega)}d\omega \\
&\qquad = \sum_{\ell\in\mathbb{Z}} \int_\ell^{\ell+1} e^{-2\pi i\omega k}|\widehat{\varphi}(\omega)|^2 d\omega \\
&\qquad = \int_0^1 e^{-2\pi i\omega k} \sum_{\ell\in\mathbb{Z}} |\widehat{\varphi}(\omega+\ell)|^2 d\omega
\end{aligned}$$

Hence φ gives rise to an ONS iff the function

$$g(\omega) = |\widehat{\varphi}(\omega + \ell)|^2 \quad ,$$

which can be viewed as an L^1-function on the interval $[0, 1]$, has the same Fourier coefficients as the constant function, and since the Fourier transform is injective on $L^1[0, 1]$ (by 1.4.2) it follows that this is the case iff $g \equiv 1$. □

We already commented on the fact that we only discuss MRA's in terms of their scaling function. The following simple but crucial lemma will allow to replace the scaling function by something yet simpler, the **scaling coefficients** (simpler because they are a discrete sequence of numbers).

Lemma 3.1.5 *(***Scaling equations***, Characterisation of* $(M1)$*)*
Let $\varphi \in L^2(\mathbb{R})$*,* $V_0 = \overline{\text{span}(T_k\varphi : k \in \mathbb{Z})}$*,* $V_j = D_{2^{-j}} V_0$*. Then* $(V_j)_{j\in\mathbb{Z}}$ *has the inclusion property* $(M1)$ *of an MRA iff there exists* $(a_k)_{k\in\mathbb{Z}} \in \ell^2(\mathbb{Z})$ *such that* φ *fulfills the* **scaling equation**

$$(S1) \qquad \varphi\left(\frac{x}{2}\right) = \sum_{k\in\mathbb{Z}} a_k \varphi(x - k) \quad ,$$

in the L^2*-sense.*

Proof. Obviously $(V_j)_{j\in\mathbb{Z}}$ fulfills $(M1)$ iff $V_0 \subset V_1$. Since $(T_{k/2}\varphi)_{k\in\mathbb{Z}}$, it is obvious that $(S1)$ is equivalent to $\varphi \in V_1$, which is clearly necessary for $V_0 \subset V_1$. It is also sufficient: If $\varphi \in V_1$, then $(T_k\varphi)_{k\in\mathbb{Z}} \subset V_1$ since V_1 is invariant under halfinteger shifts. But then $V_0 \subset V_1$. □

As an aside, we note that $a_k = \sqrt{2}\langle\varphi, D_{2^{-1}} T_k\varphi\rangle$.

3.1.6 Remark and Definition The sequence $(a_k)_{k\in\mathbb{Z}}$ is called **scaling filter** or sequence of **scaling coefficients**. We define the associated **mask** as

$$m_\varphi(\xi) = \sum_{k\in\mathbb{Z}} \frac{a_k}{2} e^{-2\pi i k\xi} \quad ,$$

which converges in $L^2([-1/2, 1/2])$. (In fact, up to normalisation m_φ is just the Fourier transform of $(a_k)_{k\in\mathbb{Z}}$.) Then we obtain the following important reformulation of $(S1)$

$$(S2) \qquad \widehat{\varphi}(\xi) = m_\varphi(\xi/2)\widehat{\varphi}(\xi/2) \quad .$$

Indeed, at least in the L^2-sense, we have

$$\begin{aligned}
\widehat{\varphi}(\omega) &= \sum_{k\in\mathbb{Z}} \frac{a_k}{\sqrt{2}} (D_{1/2} T_k\varphi)^\wedge(\omega) \\
&= \sum_{k\in\mathbb{Z}} \frac{a_k}{\sqrt{2}} (D_2(M_k\widehat{\varphi})(\omega) \\
&= \sum_{k\in\mathbb{Z}} \frac{a_k}{\sqrt{2}} e^{-2\pi i k\omega/2} \frac{1}{\sqrt{2}} \widehat{\varphi}(\omega/2) \\
&= \left(\sum_{k\in\mathbb{Z}} \frac{a_k}{2} e^{-2\pi i k\omega/2}\right) \widehat{\varphi}(\omega/2) \\
&= m_\varphi(\xi/2)\widehat{\varphi}(\xi/2)
\end{aligned}$$

Iterating $(S2)$ will turn out to yield an "ansatz" for constructing φ from m_φ via

$$\widehat{\varphi}(\xi) = \widehat{\varphi}(0) \prod_{j>0} m_\varphi(\xi/2^j) \quad . \tag{3.2}$$

□

3.1.7 Exercise Compute the scaling coefficients associated to the Shannon MRA. What is the decay order? □

The following theorem gives criteria for multiresolution analyses in the most condensed form. Note that the previous lemmas contain criteria to check $(M1)$ and $(M6)$. Note also that the continuity requirement for $\widehat{\varphi}$ is not very restrictive; it holds whenever $\varphi \in \mathrm{L}^1(\mathbb{R})$.

Theorem 3.1.8 *Let $\varphi \in \mathrm{L}^2(\mathbb{R})$, and let $(V_j)_{j\in\mathbb{Z}}$ denote the associated sequence of subspaces. Assume that $\widehat{\varphi}$ is continuous at zero. Then the following are equivalent:*

(i) Properties $(M1)$ and $(M6)$ are fulfilled, and in addition $|\widehat{\varphi}(0)| = 1$.

(ii) $(V_j)_{j\in\mathbb{Z}}$ is an MRA.

Proof. $\underline{(i) \Rightarrow (ii)}$:
Only $(M2)$ and $(M3)$ remain to be shown. Let $g \in \mathrm{L}^2(\mathbb{R})$, and assume in addition that $\mathrm{supp}(g) \subset [-R, R]$. Then

$$\begin{aligned}
\|P_j g\|_2^2 &= \sum_{m\in\mathbb{Z}} |\langle g, D_{2^{-j}} T_m \varphi\langle|^2 \\
&= \sum_{m\in\mathbb{Z}} \left| \int_{-R}^{R} g(x) \left(D_{2^{-j}} T_m \varphi\right)(x) dx \right|^2 \\
&\leq \|g\|_2^2 \sum_{m\in\mathbb{Z}} \int_{-R}^{R} |(D_{2^{-j}} T_m \varphi)(x)|^2 dx \\
&= \|g\|_2^2 \sum_{m\in\mathbb{Z}} \int_{-R}^{R} |2^{j/2} \varphi(2^j x - m)|^2 dx \\
&= \|g\|_2^2 \sum_{m\in\mathbb{Z}} \int_{-m-2^j R}^{-m+2^j R} |\varphi(x)|^2 dx \quad .
\end{aligned}$$

As soon as $2^j R < 1/2$, we have

$$\sum_{m\in\mathbb{Z}} \int_{-m-2^j R}^{-m+2^j R} |\varphi(x)|^2 dx = \int_{\mathbb{R}} |\varphi(x)|^2 \cdot \chi_{A_j}(x) dx \tag{3.3}$$

where

$$A_j = \bigcup_{m\in\mathbb{Z}} [m - 2^j R, m + 2^j R] \quad .$$

Since $\chi_{A_j} \to 0$ pointwise a.e., as $j \to -\infty$, the dominated convergence theorem yields that the right-hand side of (3.3) converges to zero for $j \to -\infty$, which proves $(M3)$ for g. For

$g \in \mathrm{L}^2(\mathbb{R})$ arbitrary and $\epsilon > 0$ pick $g_1, R > 0$ with $\|g - g_1\| < \epsilon/2, g_1 \cdot \chi_{[-R,R]} = g_1$. We just proved that $\|P_j g_1\| < \epsilon/2$ for j small enough. It follows that

$$\|P_j g\| \le \|P_j g_1\| + \|P_j(g - g_1)\| \le \epsilon/2 + \|g - g_1\| = \epsilon \quad ,$$

and $(M2)$ is proved for arbitrary g. (In functional-analytic terms: We need to show $P_j \to 0$ in the strong operator topology; since the P_j are bounded in the norm, it is sufficient to check this on a dense subspace.)
For the proof of $(M2)$ let $f \perp V_j$. Pick $g \in \mathrm{L}^2(\mathbb{R})$ with $\widehat{g} = \widehat{f} \cdot \chi_{[-R,R]}$, and $\|f - g\| < \epsilon$. By assumption, $\|P_j f\| = 0$, and thus

$$\|P_j g\| \le \|P_j f\| + \|P_j(f - g)\| \le \epsilon \quad .$$

Then we compute

$$\begin{aligned} \|P_j g\|^2 &= \sum_{m \in \mathbb{Z}} |\langle g, D_{2^{-j}} T_m \varphi \rangle|^2 \\ &= \sum_{m \in \mathbb{Z}} \left| \langle \widehat{g}, (D_{2^{-j}} T_m \varphi)^\wedge \rangle \right|^2 \\ &= \sum_{m \in \mathbb{Z}} \left| \int_{\mathbb{R}} \widehat{g}(\omega) \widehat{\varphi}(2^{-j}\omega) 2^{-j/2} e^{2\pi i m 2^{-j}\omega} d\omega \right|^2 \\ &= \sum_{m \in \mathbb{Z}} \left| \int_{-R}^{R} h_j(\omega) e_m(\omega) d\omega \right|^2 \end{aligned}$$

where

$$h_j(\omega) = \widehat{g}(\omega) \widehat{\varphi}(2^{-j}\omega) \text{ and } e_m(\omega) = 2^{-j/2} e^{2\pi i m 2^{-j}\omega} \quad .$$

As soon as

$$[-R, R] \subset [-2^{j-1}, 2^{j-1}]$$

the sequence $(e_m)_{m \in \mathbb{Z}}$ is an ONB in

$$\mathrm{L}^2([-2^{j-1}, 2^{j-1}]) \supset \mathrm{L}^2([-R, R])$$

which allows to continue

$$\begin{aligned} \sum_{m \in \mathbb{Z}} \left| \int_{-R}^{R} h_j(\omega) e_m(\omega) d\omega \right|^2 &= \int_{-R}^{R} |h_j(\omega)|^2 d\omega \\ &= \int_{-R}^{R} |\widehat{f}(\omega)|^2 |\widehat{\varphi}(2^{-j}\omega)|^2 d\omega \\ &\to \|g\|^2 |\widehat{\varphi}(0)|^2 \end{aligned}$$

since the continuity of $\widehat{\varphi}$ at zero implies $\widehat{\varphi}(2^{-j}\omega) \to \widehat{\varphi}(0)$ **uniformly** on $[-R, R]$. It follows that for j big enough

$$\|g\|^2 \le \frac{\|P_j g\|^2}{\widehat{\varphi}(0)} + \epsilon \le \epsilon\left(\frac{1}{\widehat{\varphi}(0)} + 1\right)$$

whence

$$\|f\|^2 \le \|g\|^2 + \|f - g\|^2 \le \epsilon \left(\frac{1}{\widehat{\varphi}(0)} + 2 \right)$$

for all $\epsilon > 0$. Thus $f = 0$; and $(M3)$ is established.
$(ii) \Rightarrow (i)$:
The calculation showing $(M2)$ for the converse direction shows that $\|P_j g\|^2 \to \|g\|^2 |\widehat{\varphi}(0)|^2$, while property $(M2)$ implies $\|P_j g\|^2 \to \|g\|^2$. Thus any nonzero g yields $|\widehat{\varphi}(0)| = 1$, while $(M2)$ and $(M3)$ obviously hold. □

Note that we have only used the assumptions on φ to prove $(M2)$; $(M3)$ is in fact satisfied by all sequences of subspaces obtained from the translates and dilates of a single function along a dyadic grid.

3.1.9 Example (Haar multiresolution analysis) Define

$$V_0 = \{f \in \mathrm{L}^2(\mathbb{R}) : \forall n \in \mathbb{Z} : f|_{[n,n+1[} \text{ constant } \} \quad .$$

Then, using $\varphi = \chi_{[0,1[}$, it is not hard to see that $(T_k\varphi)_{k\in\mathbb{Z}}$ is an ONB of V_0. φ induces an MRA. This can be checked directly, but we intend to use the theorem: $(M1)$ is fulfilled, since obviously

$$V_1 = \{f \in \mathrm{L}^2(\mathbb{R}) : \forall n \in \mathbb{Z} : f|_{[n/2,n/2+1/2[} \text{ constant } \} \quad .$$

We already checked $(M6)$, and $\varphi = \chi_{[0,1]}$ is in $\mathrm{L}^1(\mathbb{R})$, and thus $\widehat{\varphi}$ is continuous, with $\widehat{\varphi}(0) = \int_{\mathbb{R}} \varphi(x)dx = 1$. Thus the theorem applies to yield that $(V_j)_{j\in\mathbb{Z}}$ is indeed a multiresolution analysis, the **Haar multiresolution analysis**. The scaling equation takes here a particularly boring form:

$$\varphi(x) = \varphi(2x) + \varphi(2x - 1)$$

□

3.2 Wavelet-ONB from MRA

So far, wavelets have not made an appearance; in particular, there should be no confusion between wavelets with scaling functions. In a sense which is made precise in the next definition, wavelets inhabit the gap between scaling functions of different scales.

Definition 3.2.1 Let $(V_j)_{j\in\mathbb{Z}}$ denote an MRA. A space V_j is called **approximation space of scale** j. Denote by W_j the orthogonal complement of V_j in V_{j+1}, so that we have the orthogonal decomposition $W_j \oplus V_j = V_{j+1}$. W_j is called the **detail space of scale** j. $\psi \in W_0$ is called a **wavelet associated to** $(V_j)_{j\in\mathbb{Z}}$ if $(T_k\psi)_{k\in\mathbb{Z}}$ is an ONB of W_0. □

Theorem 3.2.2 *Let ψ be a wavelet associated to the MRA $(V_j)_{j\in\mathbb{Z}}$, and define $\psi_{j,k} = D_{2^{-j}} T_k \psi$ $(j, k \in \mathbb{Z})$.*

(a) $(\psi_{j,k})_{k\in\mathbb{Z}}$ is an ONB of W_j, for all $j \in \mathbb{Z}$.

(b) $(\psi_{j,k})_{k\in\mathbb{Z},j<j_0}$ is an ONB of V_{j_0}, for all $j_0 \in \mathbb{Z}$.

(c) $(\psi_{j,k})_{j,k\in\mathbb{Z}}$ is a wavelet ONB of $\mathrm{L}^2(\mathbb{R})$.

Proof. Concerning part (a), it is obvious that $W_j = D_{2^{-j}} W_0$, and $(\psi_{j,k})_{k\in\mathbb{Z}}$ is just the image of the ONB $(T_k\psi)_{k\in\mathbb{Z}}$ under the unitary isomorphism, thus is an ONB itself.

For part (b) note that $\langle \psi_{j,k}, \psi_{i,m} \rangle = 0$ whenever $j < i$, since $\psi_{j,k} \in V_j \subset V_{i-1}$, and $\psi_{i,m}$ is contained in the orthogonal complement of V_{i-1}. Together with part (a) this implies that

$(\psi_{j,k})_{k\in\mathbb{Z},j<j_0}$ is an ONS. Now suppose that $f \in V_{j_0}$ with $f \perp \psi_{j,k}$ for all $j < j_0$ and $k \in \mathbb{Z}$. Then in particular $f \perp \psi_{j_0-1,k}$, for all $k \in \mathbb{Z}$, and thus $f \in W_{j_0-1}^{\perp} = V_{j_0-1}$. (Here $^\perp$ refers to the orthogonal complement in V_j, and the equality follows from $V^{\perp^\perp} = V$.) Proceeding inductively, we obtain $f \in V_j$, for all $j < j_0$. But now $(M3)$ implies $f = 0$.

Now for part (c), the orthogonality has been shown in (b). Moreover, part (b) and $(M2)$ imply that the wavelet system is total, hence an ONB. □

Our next aim is to show that wavelets can associated to every MRA, and in a rather straightforward fashion, too. The following somewhat technical lemma will turn to be useful on several occasions:

Lemma 3.2.3 *Let φ be the scaling function of an MRA, with associated mask m_φ. Then the following relation holds:*

$$(S3) \qquad |m_\varphi(\omega)|^2 + |m_\varphi(\omega + 1/2)|^2 = 1 \quad (\text{ a.e. } \omega)$$

Proof. Assuming that φ is a scaling function, an application of Lemma 3.1.4 together with relation $(S2)$ implies that

$$\begin{aligned} 1 &= \sum_{\ell\in\mathbb{Z}} |\widehat{\varphi}(\omega+\ell)|^2 \\ &= \sum_{\ell\in\mathbb{Z}} \left| m_\varphi\left(\frac{\omega+\ell}{2}\right)\widehat{\varphi}\left(\frac{\omega+\ell}{2}\right)\right|^2 \\ &= \sum_{\ell\in\mathbb{Z}} \left| m_\varphi\left(\frac{\omega+2\ell}{2}\right)\widehat{\varphi}\left(\frac{\omega+2\ell}{2}\right)\right|^2 + \left| m_\varphi\left(\frac{\omega+2\ell+1}{2}\right)\widehat{\varphi}\left(\frac{\omega+2\ell+1}{2}\right)\right|^2 \\ &= \left| m_\varphi\left(\frac{\omega}{2}\right)\right|^2 \sum_{\ell\in\mathbb{Z}} \left|\widehat{\varphi}\left(\frac{\omega}{2}+\ell\right)\right|^2 + \left| m_\varphi\left(\frac{\omega+1}{2}\right)\right|^2 \sum_{\ell\in\mathbb{Z}} \left|\widehat{\varphi}\left(\frac{\omega+1}{2}+\ell\right)\right|^2 \\ &= \left| m_\varphi\left(\frac{\omega}{2}\right)\right|^2 + \left| m_\varphi\left(\frac{\omega+1}{2}\right)\right|^2 \quad . \end{aligned}$$

Here we have used that m_φ is a 1-periodic function, as well as Lemma 3.1.4 again. □

The argument for the following lemma was actually used in the derivation of the scaling equation; we will therefore not repeat it here.

Lemma 3.2.4 *Let $\varphi \in \mathrm{L}^2(\mathbb{R})$ be such that $(T_k\varphi)_{k\in\mathbb{Z}}$ is an ONS, and let $V = \overline{\mathrm{span}(T_k : k \in \mathbb{Z})}$. Then*

$$f \in V \Leftrightarrow \exists m_f \text{ 1-periodic, }, m_f \in \mathrm{L}^2([0,1]) \text{ with } \widehat{f}(\omega) = m_f(\omega)\widehat{\varphi}(\omega)$$

We can now characterise the wavelets associated to an MRA.

Proposition 3.2.5 *Let $(V_j)_{j\in\mathbb{Z}}$ be an MRA with scaling function φ. Let $\psi \in W_0$, then*

$$\psi = \sum_{k\in\mathbb{Z}} b_k (D_{2^{-1}} T_k \varphi)$$

for a suitable sequence $(b_k)_{k\in\mathbb{Z}} \in \ell^2(\mathbb{Z})$. Defining

$$m_\psi(\omega) = \sum_{k\in\mathbb{Z}} \frac{b_k}{\sqrt{2}} e^{-2\pi i k\omega} \quad .$$

Then ψ is a wavelet associated to $(V_j)_{j\in\mathbb{Z}}$ iff the orthogonality relations hold almost everywhere

$$(W1) \qquad |m_\psi(\omega)|^2 + |m_\psi(\omega+1/2)|^2 = 1$$

$$(W2) \qquad m_\psi(\omega)\overline{m_\varphi(\omega)} + m_\psi(\omega+1/2)\overline{m_\varphi(\omega+1/2)} = 0$$

Proof. The same calculation as in the proof of $(S2)$ shows that

$$\widehat{\psi}(\xi) = m_\psi(\xi/2)\widehat{\varphi}(\xi/2) \quad .$$

Hence the same argument as the one proving 3.2.3 can be used for $(W1)$. On the other hand, since $T_k\psi \in W_0 \perp V_0$, we have

$$\begin{aligned}
0 &= \langle T_k\psi, \varphi\rangle \\
&= \int_{\mathbb{R}} e^{-2\pi i k\omega}\widehat{\psi}(\omega)\overline{\widehat{\varphi}(\omega)}d\omega \\
&= \sum_{\ell\in\mathbb{Z}} \int_\ell^{\ell+1} e^{-2\pi i k\omega}\widehat{\psi}(\omega)\overline{\widehat{\varphi}(\omega)}d\omega \\
&= \int_0^1 e^{-2\pi i k\omega} \sum_{\ell\in\mathbb{Z}} \widehat{\psi}(\omega+\ell)\overline{\widehat{\varphi}(\omega+\ell)}d\omega \\
&= \int_0^1 e^{-2\pi i k\omega} \sum_{\ell\in\mathbb{Z}} m_\psi\left(\frac{\omega+\ell}{2}\right)\overline{m_\varphi\left(\frac{\omega+\ell}{2}\right)}\left|\widehat{\varphi}\left(\frac{\omega+\ell}{2}\right)\right|^2 d\omega \\
&= \int_0^1 e^{-2\pi i k\omega} H(\omega)d\omega
\end{aligned}$$

where

$$\begin{aligned}
H(\omega) &= \sum_{\ell\in\mathbb{Z}} m_\psi\left(\frac{\omega+\ell}{2}\right)\overline{m_\varphi\left(\frac{\omega+\ell}{2}\right)}\left|\widehat{\varphi}\left(\frac{\omega+\ell}{2}\right)\right|^2 \\
&= \sum_{\ell\in\mathbb{Z}} m_\psi\left(\frac{\omega+2\ell}{2}\right)\overline{m_\varphi\left(\frac{\omega+2\ell}{2}\right)}\left|\widehat{\varphi}\left(\frac{\omega+2\ell}{2}\right)\right|^2 \\
&\quad + m_\psi\left(\frac{\omega+2\ell+1}{2}\right)\overline{m_\varphi\left(\frac{\omega+2\ell+1}{2}\right)}\left|\widehat{\varphi}\left(\frac{\omega+2\ell+1}{2}\right)\right|^2 \\
&= m_\psi\left(\frac{\omega}{2}\right)\overline{m_\varphi\left(\frac{\omega}{2}\right)}\sum_{\ell\in\mathbb{Z}}\left|\widehat{\varphi}\left(\frac{\omega}{2}+\ell\right)\right|^2 \\
&\quad + m_\psi\left(\frac{\omega}{2}+1/2\right)\overline{m_\varphi\left(\frac{\omega}{2}+1/2\right)}\sum_{\ell\in\mathbb{Z}}\left|\widehat{\varphi}\left(\frac{\omega+1}{2}+\ell\right)\right|^2 \\
&= m_\psi\left(\frac{\omega}{2}\right)\overline{m_\varphi\left(\frac{\omega}{2}\right)} + m_\psi\left(\frac{\omega}{2}+1/2\right)\overline{m_\varphi\left(\frac{\omega}{2}+1/2\right)} \quad .
\end{aligned}$$

Applying uniqueness of Fourier series to the function H, we conclude $(W2)$. Hence we have in fact shown that

$$(W1),(W2) \Leftrightarrow (T_k\psi)_{k\in\mathbb{Z}} \cup (T_k\varphi)_{k\in\mathbb{Z}} \subset V_1 \text{ ONS} \quad .$$

We show next that the right hand side is even an ONB of V_1. For this purpose observe that $(S3), (W1), (W2)$ can be summarised as

$$(m_\varphi(\xi), m_\varphi(\xi+1/2))^t, (m_\psi(\xi), m_\psi(\xi+1/2))^t \subset \mathbb{C}^2 \text{ is an ONB.} \tag{3.4}$$

Now suppose $f \in V_1$ is orthogonal to all $T_k\varphi$ and $T_k\psi$. Writing

$$\widehat{f}(\xi) = m_f(\xi/2)\widehat{\varphi}(\xi/2)$$

for a suitable 1-periodic L^2-function, the same argument as the one proving $(W2)$ shows that almost everywhere

$$m_\psi(\omega)\overline{m_f(\omega)} + m_\psi(\omega+1/2)\overline{m_f(\omega+1/2)} = 0$$

and

$$m_\varphi(\omega)\overline{m_f(\omega)} + m_\varphi(\omega+1/2)\overline{m_f(\omega+1/2)} = 0 \quad .$$

Now these two relations amount to saying that the scalar product of $(m_f(\xi), m_f(\xi+1/2))^t$ with the two basis vectors from (3.4) vanishes. Thus $(m_f(\xi), m_f(\xi+1/2))^t = 0$, almost everywhere, and thus $f = 0$. Therefore the system $(T_k\psi)_{k\in\mathbb{Z}} \cup (T_k\varphi)_{k\in\mathbb{Z}}$ is total in V_1. But then $(T_k\psi)_{k\in\mathbb{Z}}$ is total in W_0, for if $f \in W_0$ then $f \perp T_k\varphi$ for all $k \in \mathbb{Z}$. If in addition $f \perp T_k\psi$ for all k, then $f = 0$. □

Theorem 3.2.6 *(Existence theorem for wavelets)*
Let $\varphi \in \mathrm{L}^2(\mathbb{R})$ denote an MRA of $\mathrm{L}^2(\mathbb{R})$, with scaling coefficients $(a_k)_{k\in\mathbb{Z}}$. Then

$$\psi(x) = \sum_{n\in\mathbb{Z}} (-1)^{1-n}\overline{a_{1-n}}\varphi(2x-n) \tag{3.5}$$

defines a wavelet associated to the MRA.

Proof. We first compute $m_\psi(\omega)$,

$$\begin{aligned}
m_\psi(\omega) &= \sum_{n\in\mathbb{Z}} \overline{a_{1-n}}(-1)^{1-n}e^{-2\pi i n\omega} \\
&= e^{-2\pi i\omega}\sum_{n\in\mathbb{Z}} \overline{a_{1-n}}(-1)^{1-n}e^{-2\pi i(n-1)\omega} \\
&= e^{-2\pi i\omega}\overline{\sum_{n\in\mathbb{Z}} a_{1-n}(-1)^{1-n}e^{-2\pi i(1-n)\omega}} \\
&= e^{-2\pi i\omega}\overline{\sum_{n\in\mathbb{Z}} a_{1-n}e^{-2\pi i(1-n)(\omega+1/2)}} \\
&= e^{-2\pi i\omega}\overline{\sum_{n\in\mathbb{Z}} a_n e^{-2\pi i n(\omega+1/2)}} \\
&= e^{-2\pi i\omega}\overline{m_\varphi(\omega+1/2)} \quad .
\end{aligned}$$

This implies

$$|m_\psi(\omega)|^2 + |m_\psi(\omega+1/2)|^2 = |m_\varphi(\omega+1/2)|^2 + |m_\varphi(\omega+1)|^2 = 1 \quad ,$$

e.g., $(W1)$. In addition

$$\begin{aligned} & m_\psi(\omega)\overline{m_\varphi(\omega)} + m_\psi(\omega+1/2)\overline{m_\varphi(\omega+1/2)} \\ = & \; e^{-2\pi i\omega}\overline{m_\varphi(\omega+1/2)m_\varphi(\omega)} + e^{-2\pi i(\omega+1/2}\overline{m_\varphi(\omega+1)m_\varphi(\omega+1/2)} \\ = & \; e^{-2\pi i\omega}(\overline{m_\varphi(\omega+1/2)m_\varphi(\omega)} - \overline{m_\varphi(\omega+1)m_\varphi(\omega+1/2)}) \\ = & \; 0 \;\; , \end{aligned}$$

which is $(W2)$. Thus 3.2.5 implies that ψ is a wavelet. □

The choice of ψ seems somewhat arbitrary. However, there are several desirable and nontrivial properties that ψ (as constructed in the theorem) inherits from φ, and which are not trivial, such like real values, or compact support: If φ has compact support, it is obvious that the scaling sequence is finite. Thus ψ is a finite linear combination of shifts of a function with compact support, thus also has compact support. In general, if we do not postulate additional properties, neither the scaling functions nor the wavelet associated to an MRA are uniquely given. The following example makes this precise for the Shannon-MRA. We will see below however that requiring compact support does make the wavelet and scaling function unique up to integer shifts.

3.2.7 Exercise (Shannon-MRA revisited) Let $(V_j)_{j\in\mathbb{Z}}$ be the Shannon MRA.

(a) Prove that $\widetilde{\varphi}$ is a scaling function associated to $(V_j)_{j\in\mathbb{Z}}$ iff $|\widehat{\varphi}| = \chi_{[-1/4,1/4]}$.

(b) Prove that $\psi \in \mathrm{L}^2(\mathbb{R})$ is a wavelet associated to the MRA iff $|\psi(\omega)| = \chi_{[-1/2,-1/4]} + \chi[1/4, 1/2]$.

(c) Prove that $(a_k)_{k\in\mathbb{Z}}$ is the sequence of scaling coefficients for some scaling function of the MRA iff m_φ fulfills

$$|m_\varphi(\omega)| = \chi_{[-1/4,1/4]} \quad .$$

□

3.3 Wavelet-ONB's with compact support

The aim of this section is the construction of MRA's such that the associated wavelets have additional properties such as compact support, smoothness and decay.

Remark 3.3.1 If the scaling function φ associated to an MRA has compact support, then in particular $\langle \varphi, D_{2^{-1}}T_k\varphi\rangle = 0$ for large k. Hence the associated scaling sequence has compact support also, as does the wavelet associated to φ by 3.2.6.

It is therefore a reasonable approach to start from a finite sequence of scaling coefficients (a_k), such that a solution φ of the associated scaling equation exists and induces an MRA.

For the construction of φ we use the "ansatz" already sketched in Remark 3.1.6, i.e., iterating the scaling equation on the Fourier transform side. □

Theorem 3.3.2 *Let $(a_k)_{k=T,\ldots,S}$ be a finite sequence of complex numbers, define $m(\omega) = \sum_{k=T}^{S} \frac{a_k}{2} e^{-2\pi i k\omega}$. Assume that m fulfills*

1. $|m(\omega)|^2 + |m(\omega+1/2)|^2 = 1$ for all $\omega \in \mathbb{R}$.

2. $m(0) = 1$.

3. $m(\omega) \neq 0$ *for all* $\omega \in [-1/4, 1/4]$.

Define

$$\Phi(\omega) = \prod_{j>1} m(2^{-j}\omega) \quad . \tag{3.6}$$

Then the product converges uniformly on compacta, yielding a continuous function Φ. $\varphi = \mathcal{F}^{-1}(\Phi)$ *is the scaling function of an MRA, with* $\mathrm{supp}(\varphi) \subset [S, T]$.

Remark 3.3.3 Property 1. is just the relation $(S1)$, and thus necessary by 3.2.3. Property 2. is obviously necessary for the product (3.6) to converge at $\omega = 0$. Hence the only new and not strictly necessary condition is 3. In fact a certain weakening of 3., known as Cohen's condition, also ensures convergence; confer [6, 5] for details. □

Proof. We need to show

(i) Φ is well-defined and in $\mathrm{L}^2(\mathbb{R})$.

(ii) $\mathcal{F}^{-1}(\varphi)$ fulfills the conditions of Theorem 3.1.8, and $\mathrm{supp}(\varphi) \subset [S, T]$.

We start the proof of (i) with pointwise convergence. Condition 1. implies in particular $|m(\omega)| \leq 1$. Moreover the trigonometric polynomial m is continuously differentiable, hence there exists a constant $C > 0$ with $|m(\xi) - 1| \leq C|\xi|$. (Here we used $m(\xi) = 1$.) Hence $|m(2^{-j}\xi) - 1| \leq C2^{-j}|\xi|$, and thus

$$\left| \prod_{j=1}^{j_0+1} m(2^{-j}\xi) - \prod_{j=1}^{j_0} m(2^{-j}\xi) \right| \leq \left| \prod_{j=1}^{j_0} m(2^{-j}\xi) \right| C2^{-j_0}|\xi| \leq C2^{-j_0}|\xi| \quad .$$

Hence a simple telescope sum argument involving the geometric series yields existence of

$$\lim_{j_0 \to 1} \prod_{j=1}^{j_0} m(2^{-j}\xi) \quad ,$$

in fact, we obtain uniform convergence on every bounded set ($|\xi| \leq R$ yields uniform estimates in $[-R, R]$). Hence Φ is a locally uniform limit of continuous functions, thus continuous itself. $m(0) = 1$ yields $\Phi(0) = 1$. By construction, $\Phi(\xi) = m(\xi/2)\Phi(\xi/2)$, which will account for property $(M1)$ of an MRA once we have shown that $\Phi \in \mathrm{L}^2(\mathbb{R})$. Hence it remains to prove $\Phi \in \mathrm{L}^2(\mathbb{R})$, and property $(M6)$. For this purpose we introduce the auxiliary functions

$$\Phi_N(\xi) = \prod_{j=1}^{N} m(2^{-j}\xi) \cdot \chi_{[-2^{N-1}, 2^{N-1}]}$$

and the quantities

$$I_N^k = \int_{-2^{N-1}}^{2^{N-1}} |\Phi_N(\xi)|^2 e^{2\pi i k \xi} d\xi$$

In the following computations we will repeatedly use the fact that

$$\omega \mapsto \prod_{j=1}^{N} m(2^{-j}\xi)$$

is periodic with period 2^N. We first prove

$$(*) \qquad I_N^k = \delta_{0,k} \quad , \forall N \in \mathbb{N}, \forall k \in \mathbb{Z} \quad .$$

Indeed,

$$\begin{aligned} I_1^k &= \int_{-1}^{1} |m(\xi/2)|^2 e^{-2\pi i k\xi} d\xi \\ &= \int_{-1}^{0} \left(|m(\xi/2)|^2 + |m(\xi/2+1/2)|^2\right) e^{-2\pi i k\xi} d\xi \\ &= \delta_{0,k} \quad , \end{aligned}$$

using property 1. Now assume $I_N^k = \delta_{0,k}$ for $N \geq 1$. Then

$$\begin{aligned} I_{n+1}^k &= \int_{-2^N}^{2^N} \left| \prod_{j=1}^{N+1} m(2^{-j}\xi) \right|^2 e^{-2\pi k\xi} d\xi \\ &= \int_{-2^N}^{2^N} \left|m(2^{-N-1}\xi)\right|^2 \left| \prod_{j=1}^{N} m(2^{-j}\xi) \right|^2 e^{-2\pi k\xi} d\xi \\ &= \int_{-2^N}^{0} \left(\left|m(2^{-N-1}\xi)\right|^2 + \left|m(2^{-N-1}\xi + 1/2)\right|^2 \right) \left| \prod_{j=1}^{N} m(2^{-j}\xi) \right|^2 e^{-2\pi k\xi} d\xi \\ &\overset{1.}{=} \int_{-2^N}^{0} \left| \prod_{j=1}^{N} m(2^{-j}\xi) \right|^2 e^{-2\pi k\xi} d\xi \\ &= \int_{-2^{N-1}}^{2^{N-1}} \left| \prod_{j=1}^{N} m(2^{-j}\xi) \right|^2 e^{-2\pi k\xi} d\xi \\ &= I_N^k = \delta_{k,0} \quad , \end{aligned}$$

using the induction hypothesis. Now $(*)$ and $|m(\xi)| \leq 1$ imply in particular for all $N \geq 1$ that

$$\int_{-2^{N-1}}^{2^{N-1}} |\Phi(\xi)|^2 d\xi \leq \int_{-2^{N-1}}^{2^{N-1}} |\Phi_N(\xi)|^2 d\xi \leq I_N^0 = 1 \quad ,$$

in particular $\Phi \in \mathrm{L}^2(\mathbb{R})$, with $\|\Phi\|_2^2 \leq 1$. Next we need to prove property $(M6)$ for Φ, which by 3.1.4 is the same as

$$\sum_{\ell \in \mathbb{Z}} |\Phi(\xi + \ell)|^2 = 1 \quad .$$

Using the Fourier transform on the interval, this amounts to proving

$$(**) \qquad \delta_{0,k} = \int_{0,k} \sum_{\ell\in\mathbb{Z}} |\Phi(\xi+\ell)|^2 e^{-2\pi i k\xi} d\xi$$
$$= \int_{\mathbb{R}} \|\Phi(\xi)|^2 e^{-2\pi i k\xi} d\xi \quad .$$

In view of $(*)$ it is enough to prove

$$(***) \qquad \lim_{N\to\infty} I_N^k = \int_{\mathbb{R}} |\phi(\xi)|^2 e^{-2\pi k\xi} d\xi \quad .$$

By definition of Φ and Φ_N we obtain for $\xi \in [-2^{N-1}, 2^{N-1}]$ the equation

$$\Phi_N(\xi)\Phi(2^{-N}\xi) = \Phi(\xi) \quad . \tag{3.7}$$

Our aim is to estimate $|\Phi_N|^2$ from above by a single L^1-function (using (3.7), and then use the pointwise convergence $\Phi_N \to \Phi$ to apply the dominated convergence theorem for the proof of $(***)$. For the upper estimate pick a constant $\alpha > 0$ with $||m(\xi)|^2 - 1| \leq \alpha|\xi|$ on $[-1/2, 1/2]$; note that $|m(\xi)|^2$ is continuously differentiable. This entails

$$|m(\xi)|^2 \geq |1 - \alpha|\xi|| \quad .$$

Moreover, assumption 3. together with continuity of m implies in fact

$$|m(\xi)| \geq c > 0$$

on $[-1/4, 1/4]$. Pick j_0 such that $2^{-j_0-1}\alpha < 1$. Then, for all $\xi \in [-1/2, 1/2]$,

$$\begin{aligned}
|\Phi(\xi)|^2 &= \prod_{j=1}^{j_0} |m(2^{-j}\xi|^2 \cdot \prod_{j=j_0+1}^{\infty} |m(2^{-j}\xi|^2 \\
&\geq c^{2j_0} \prod_{j=j_0}^{\infty} |1 - 2^{-j-1}\alpha|^2 \\
&= c^{2j_0} \prod_{j=1}^{\infty} |1 - 2^{-j_0-1}\alpha \\
&\geq c^{2j_0} \prod_{j=1}^{\infty} |1 - 2^{-j}|^2 \\
&\geq c^{2j_0} C
\end{aligned}$$

with a constant $C > 0$ (see next lemma). In any case, we have produced a strictly positive lower bound C_0 for Φ on $[-1/2, 1/2]$. Therefore, (3.7) shows

$$|\Phi_N(\xi)| \leq C_0 |\Phi(\xi)| \quad ,$$

and we have finally proved the integrable upper bound for the $|\Phi_N|^2$. Thus $(***)$ is proved, which in turn proves $(M6)$.

The last thing to show is that $\text{supp}(\phi) \subset [S, T]$. This requires the Paley-Wiener theorem which characterises Fourier transforms of compactly supported functions, and we refer the reader to [16, Lemma 4.3] for this additional argument. □

Lemma 3.3.4 $\prod_{j=1}^{\infty} \frac{2^j-1}{2^j} > 0$.

Proof. We have

$$\log\left(\prod_{j=1}^{N} \frac{2^j-1}{2^j}\right) = \sum_{j=1}^{n} \log(1-2^{-j}) \quad .$$

Then for $x \in [0,1)$

$$\begin{aligned}
\log(1-x) &= -\int_0^x \frac{1}{1-t} dt \\
&= -\int_0^x \sum_{k=0}^{\infty} t^k dt \\
&= -\sum_{k=0}^{\infty} \frac{x^{k+1}}{k+1} \\
&= -\sum_{k=1}^{\infty} \frac{x^k}{k} \quad .
\end{aligned}$$

Hence

$$\log\left(\prod_{j=1}^{N} \frac{2^j-1}{2^j}\right) = -\sum_{j=1}^{N}\sum_{k=1}^{\infty} \frac{1}{k2^jk} \quad .$$

Computing

$$\begin{aligned}
-\sum_{j=1}^{\infty}\sum_{k=1}^{\infty} \frac{1}{k2^jk} &= -\sum_{k=1}^{\infty} \frac{1}{k} \sum_{j=1}^{\infty} (2^{-k})^j \\
&= -\sum_{k=1}^{\infty} \frac{1}{k}\left(\frac{1}{1-2^{-k}} - 1\right) \\
&= -\sum_{k=1}^{\infty} \frac{1}{k(2^k-1)} > -\infty \quad .
\end{aligned}$$

Hence the sum of logs converges absolutely, and thus $\prod_{j=1}^{\infty} \frac{2^j-1}{2^j} > 0$. □

We already mentioned **trigonometric polynomials**, i.e., functions of the form

$$m(\omega) = \sum_{k=S}^{T} a_k e^{-2\pi i k\omega} = p(e^{-2\pi i\omega}) \quad .$$

Here we can write $p(z) = z^{-S}q(z)$, and q is the polynomial

$$q(z) = \sum_{k=0}^{T-S} a_{k+S} z^k \quad .$$

In particular we may use factorisation properties of polynomials. This observation is used in the next proposition, which addresses the problem of ensuring enough vanishing moments in the construction procedure.

Proposition 3.3.5 *Let φ be a scaling function with compact support. Then*

$$m_\varphi(\omega) = (e^{-2\pi i\omega} - 1)^\ell \widetilde{m}(\omega)$$

for an integer $\ell \geq 1$ and a trigonometric polynomial. Then, if ψ is the wavelet associated to φ by Theorem 3.2.6, then ψ has ℓ vanishing moments.

Proof. m_φ is a trigonometric polynomial, hence $(S2)$, $|m_\varphi(\omega)|^2 + |m_\varphi(\omega + 1/2)|^2 + 1$ holds everywhere on $\mathbb{R}$. In particular, $|m_\phi(0)|^2 = 1$ implies $m_\phi(1/2) = 0$. This implies the factorisation $m_\varphi(\omega) = (e^{-2\pi i\omega} - 1)^\ell \widetilde{m}(\omega)$, with $\ell \geq 1$. For the associated wavelet we have

$$\begin{aligned} \widehat{\psi}(\omega) &= m_\psi(\omega/2)\widehat{\varphi}(\omega/2) \\ &= \overline{m_\varphi(\omega/2 + 1/2)} e^{-2\pi i\omega}\widehat{\varphi}(\omega/2) \quad , \end{aligned}$$

compare the proof of 3.2.6. Plugging in the factorisation of m_φ, we obtain

$$\widehat{\psi}(\omega) = \left(1 - e^{\pi i\omega}\right)^\ell \overline{\widetilde{m}(\omega/2 + 1/2)}\widehat{\varphi}(\omega/2) \quad .$$

The first factor yields a zero of order ℓ at zero. Since $\mathrm{supp}(\psi)$ is compact, we have $t \mapsto t^n\psi(t) \in \mathrm{L}^1(\mathbb{R})$. Thus we may apply 1.6.5 to obtain that the n-th moment vanishes, for $0 \leq n < \ell$. □

The following property can be seen as an interpolation property of MRA with compact supports and vanishing moments:

Proposition 3.3.6 *Let φ be a scaling function with compact support, such that the associated wavelet has ℓ vanishing moments. Define*

$$c_{m,k} = \int_{\mathbb{R}} x^m \overline{\varphi(x-k)} dx \quad ,$$

for all $0 \leq m < \ell$ and all $k \in \mathbb{Z}$. Then, for almost all $x \in \mathbb{R}$,

$$\sum_{k\in\mathbb{Z}} c_{m,k}\varphi(x-k) = x^m \tag{3.8}$$

Note that since $\mathrm{supp}(\varphi)$ is compact, the sum on the left hand side is finite for all x. Convergence is thus not a problem. Moreover, it is obvious that the equality holds for all $x \in \mathbb{R}$ if φ is continuous.

Proof. By construction of the W_j we have

$$\mathrm{L}^2(\mathbb{R}) = V_0 \oplus \bigoplus_{j\geq 1} W_j$$

and $(T_k\varphi)_{k\in\mathbb{Z}} \bigcup (D_{2^{-j}}T_k\psi)_{j\geq 1, k\in\mathbb{Z}}$ is an ONB of $\mathrm{L}^2(\mathbb{R})$. Fix $R > 0$ and $N \in \mathbb{N}$ such that $\mathrm{supp}(\psi), \mathrm{supp}(\varphi) \subset [-N, N]$. Define $f(x) = x^m \chi_{-R-2N, R+2N}$. Then

$$f(x) = \sum_{k\in\mathbb{Z}} \widetilde{c_k}\varphi(x-k) + \sum_{j\geq 1, k\in\mathbb{Z}} d_{j,k}\psi_{j,k}(x) \tag{3.9}$$

almost everywhere on $\mathbb{R}$. Here the $\widetilde{c_k}$ and $d_{j,k}$ are the expansion coefficients of f with respect to the ONB. Now pick $x \in [-R, R]$. If $(j,k) \in \mathbb{N} \times \mathbb{Z}$ is such that $x \in \text{supp}(\psi_{j,k})$, it follows that $\text{supp}(\psi_{j,k}) \subset [-2N - R, 2N + R]$, since

$$\text{supp}(\psi_{j,k}) \subset [2^{-j}(-N + k), 2^{-j}(N + k)]$$

which has length less than N, and $\text{supp}(\psi_{j,k})$ is assumed to contain x. Hence we obtain

$$\begin{aligned} d_{j,k} &= \int_{-N-R}^{N+R} x^m \overline{\psi_{j,k}(x)} dx \\ &= \int_{\mathbb{R}} x^m \overline{\psi_{j,k}(x)} dx \\ &= \int_{\mathbb{R}} x^m 2^{j/2} \overline{\psi(2^j x - k)} dx \\ &= \int_{\mathbb{R}} 2^{-j/2} \left(\frac{x-k}{2^j} \right)^m \overline{\psi}(x) dx \\ &= 0 \ , \end{aligned}$$

since a polynomial of order m is a linear combination of monomials of order up to m, which is also not "seen" by the wavelet. Now a similar argument shows that $\widetilde{c_k} = c_{m,k}$ for all k with $x \in \text{supp}(T_k\varphi)$. Thus (3.8) and (3.9) coincide on $[-R, R]$. Letting $R \to \infty$ yields the full statement. □

The following proposition settles a uniqueness issue. If we start with a finite scaling sequence and construct a compactly supported scaling function and wavelet from it by the procedure from Theorem 3.3.2, then these functions are unique up to shifts.

Proposition 3.3.7 *Let $\varphi, \widetilde{\varphi} \in \mathrm{L}^2(\mathbb{R})$ be compactly supported. Assume that $(T_k\varphi)_{k\in\mathbb{Z}}$ and $(T_k\widetilde{\varphi})_{k\in\mathbb{Z}}$ are ONB's of the same closed subspace $V \subset \mathrm{L}^2(\mathbb{R})$. Then there exists $\ell \in \mathbb{Z}$ and $c \in \mathbb{C}$ with $|c| = 1$ such that $\phi = cT_\ell\widetilde{\varphi}$.*

Proof. Exercise. (Hint: From 3.2.4 we know $\widehat{\varphi}\widehat{\widetilde{\varphi}} \cdot m_1$, and $\widehat{\widetilde{\varphi}} = \widehat{\varphi}$. Here m_1,m_2 are trigonometric polynomials, since $\varphi, \widetilde{\varphi}$ have compact supports. It follows that $m_1 m_2 = 1$, and this implies the desired statement.) □

The following theorem is the most important result of this chapter: There exist compactly supported wavelet ONB's with arbitrary smoothness and number of vanishing moments. In particular, part (c) of the theorem implies that one has to increase the discrete filter size roughly by 10 in order to obtain an increase in regularity by one differentiation order.

Theorem 3.3.8 *(Daubechies)*
There exists a family of finite scaling sequences $a_N = (a_k, N)_{k\in\mathbb{Z}}$, $N \in \mathbb{N}$ with the following properties

- *(a) The solution φ_N of the associated scaling equation is a real valued scaling function of an MRA.*
- *(b)* $\text{supp}(a_N) \subset \{0, \ldots, 2N-1\}$, $\text{supp}(\varphi_N) \subset [0, 2N-1]$ *and* $\text{supp}(\psi_N) \subset [1-N, N]$, *where ψ_N is chosen according to Theorem 3.2.6.*
- *(c) ψ_N has N vanishing moments.*

(d) $\varphi_N, \psi_N \in C^{\lfloor \alpha N - \epsilon \rfloor}$, *for all* $\epsilon > 0$ *and all* $N > N(\epsilon)$. *Here* $\alpha = 1 - \frac{\ln 3}{2 \ln 2} \approx 0.20775$.

Proof. (Only a sketch. Confer [3, 16, 5] for details.)
Step 1. Define the function

$$M_N(\omega) = 1 - c_N \int_0^\omega \sin^{2N-1}(2\pi t) dt$$

with $\frac{1}{c_N} = \int_0^1 \sin^{2N-1}(2\pi t)dt$. Then M_N is a trigonometric polynomial of degree $4N$ which fulfills the relations

$$M_N(\omega) + M_N(\omega + 1/2) = 1 \quad , \quad M_N(\omega) \geq 0 \quad .$$

Step 2. By a theorem due to Szegö there exists a trigonometric polynomial

$$m_N(\omega) = \sum_{k=0}^{2N-1} \frac{a_{k,N}}{2} e^{-2\pi i \omega k}$$

with $|m_N|^2 = M_N$. (Caution: m_N is not unique, and taking $\sqrt{M_N}$ usually does not work; this function need not be a trigonometric polynomial.)
Step 3. Properties (1) and (2) from Theorem 3.3.2 are guaranteed by the construction. (3) (or the weaker yet sufficient "Cohen's criterion") follow from additional considerations.
Step 4. Smoothness properties follow from 1.6.1 and estimates of the decay of $\widehat{\varphi}$. □

3.3.9 Examples The first three filters are given by

N	a_0	a_1	a_2	a_3	a_4	a_5	a_6
1	1	1					
2	$\frac{1+\sqrt{3}}{4}$	$1\frac{3+\sqrt{3}}{4}$	$\frac{3-\sqrt{3}}{4}$	$\frac{1-\sqrt{3}}{4}$			
3	0.0498	−0.1208	−0.1909	0.6504	1.1411	0.4705	

□

Remark 3.3.10 For $N = 1$ we obtain the scaling function $\varphi_1 = \chi_{[0,1]}$ and the Haar wavelet. At the other end of the scale, it can be shown that $\lim_{N\to\infty} \widehat{\varphi_N}(\omega) = \chi_{[-1/2,1/2]}$, i.e., in a sense we obtain the Shannon-MRA as the limit case. □

3.3.11 Plotting scaling functions and wavelets. For most orthonormal wavelets there does not exist a closed analytic expression; plotting the graph of such a function is therefore something of a challenge. A possible way out consists in regarding the scaling equation

$$\varphi(x) = \sum_{k \in \mathbb{Z}} a_k \varphi(2x - k)$$

as a fixed point equation. The standard approach to the solution of such fixed point equations consists in iteration. Hence we might want to pick

$$\varphi_0(x) = \begin{cases} 1 & x = 0 \\ 0 & |x| \geq 1 \end{cases}$$

and linearly interpolating in between, and then define

$$\varphi_{n+1}(x) = \sum_{k\in\mathbb{Z}} a_k \varphi_n(2x-k) \quad .$$

It can be shown that this converges pointwise to φ, at least for the Daubechies wavelets (see [12] for details). (Note that at least for the Haar wavelet this convergence can not be uniform.) Moreover the φ_n are easily computed: It is obvious that φ_n is piecewise affine with the nodes contained in $2^{-n}\mathbb{Z}$. It thus suffices to know the values $g_{n,k} = \varphi_n(2^{-n}k)$. But these obey the recursion formula

$$\begin{aligned} g_{0,k} &= \delta_{0,k} \\ g_{n+1,\ell} &= \sum_{k\in\mathbb{Z}} g_{n,\ell-2^n k} a_k \quad . \end{aligned}$$

This recursion is implemented in the matlab wavelet toolbox under the name wavefun.

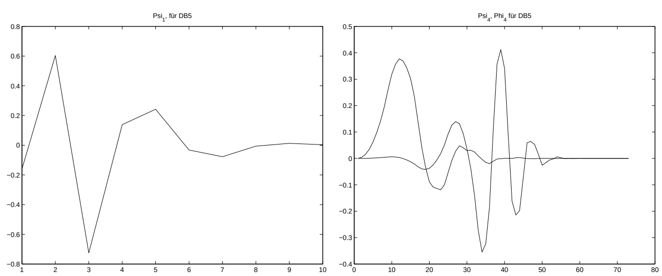

Figure 3.1: Plots of the fifth Daubechies wavelet obtained by one and four iterations. The four iterations plot contains both the approximated scaling function and the wavelet. Four iterations already yield a rather good approximation.

□

Remark 3.3.12 In the course of this script three different normalisations of the scaling coefficients have been used. Letting $a_k = \sqrt{2}\langle \varphi, D_{2^{-1}} T_k \varphi\rangle$, we used the relations

$$\begin{aligned} \varphi(x) &= \sum_{k\in\mathbb{Z}} a_k \varphi(2x-k) \\ &= \sum_{k\in\mathbb{Z}} \frac{a_k}{\sqrt{2}} \left(D_{2^{-1}} T_k \varphi\right)(x) \end{aligned}$$

as well as in the definition of m_φ

$$m_\varphi(\omega) = \sum_{k\in\mathbb{Z}} \frac{a_k}{2} e^{-2\pi i k\omega} \quad .$$

The different normalisations are the most convenient choices for their respective contexts. The differences are only partly the result of the author's sloppiness; in fact all normalisations can be found in the literature. But this implies that in order to use filters given in the literature one needs to know to which convention the source at hand refers. For this purpose the relation

$$1 = |m_\varphi(0)| = \left| \sum_k \frac{a_k}{2} \right|$$

is useful. In particular, if some filter coefficients $(\widetilde{a_k})_{k\in\mathbb{Z}}$ are found in the literature, then letting

$$a_k = \frac{2}{|\sum_k \widetilde{a}_k|}\widetilde{a_k}$$

yields the normalisation used in this script.

In particular, calling the matlab routine wfilters yields the scaling coefficients $(\widetilde{a_k})_{k\in\mathbb{Z}}$, in the normalisation $\widetilde{a_k} = \frac{a_k}{\sqrt{2}}$. □

3.4 Spline Wavelets

Daubechies' construction starts from a clever choice of scaling coefficients, constructs the scaling function and finally the MRA. This way we are guaranteed finite scaling filters (the use of that will become clear in the next chapter) but we do not really have a good understanding of the spaces V_j associated to the scaling function (we don't even know their basic building blocks explicitly). An alternative way might consist in prescribing the spaces $(V_j)_{j\in\mathbb{Z}}$ in a suitable way, and then obtain the scaling coefficients from the inclusion property. In this section we want to sketch such a construction for a particular family of spaces, namely spline spaces.

Definition 3.4.1 Let $a > 0$ and $m \in \mathbb{N}$. Define

$$S^m(a\mathbb{Z}) := \{f : \mathbb{R} \to \mathbb{C} : f \in C^{m-1}(\mathbb{R}), f|_{[ak,a(k+1)[} \text{ polynomial of degree } \le m\} \quad .$$

$f \in S^m(a\mathbb{Z})$ is called **spline of order** m **with nodes in** $a\mathbb{Z}$. □

The aim of this section is to sketch a proof of the following theorem:

Theorem 3.4.2 *Fix $m \in \mathbb{N}$, and let for $j \in \mathbb{Z}$*

$$V_j = S^m(2^{-j}\mathbb{Z}) \cap \mathrm{L}^2(\mathbb{R}) \quad .$$

Then $(V_j)_{j\in\mathbb{Z}}$ is an MRA. There exists a scaling function φ and a wavelet ψ which fulfill the following estimates

$$|\psi(x)| \le Ce^{-\alpha_1|x|} \quad , \quad |\varphi(x)| \le C^{-\alpha_2|x|} \quad ,$$

and the associated scaling coefficients satisfy

$$|a_k| \le Ce^{-\alpha_3|k|} \quad ,$$

with suitable $\alpha_i > 0$.

The following definition will not give the scaling functions, but basic building blocks which are fairly close to the scaling functions:

Definition 3.4.3 (B-splines)
Define N_m (for $m \ge 0$) recursively by $N_0 = \chi_{[0,1]}$ and $N_{m+1} = N_0 * N_m$. □

With Lemma 3.1.4 in mind, the following lemma can be read as a description how far the translates of a B-spline are from being an ONB.

Lemma 3.4.4 *For all $m \in \mathbb{N}$ there exist constants $0 < C_1 \le C_2 < \infty$ such that for all ω*

$$C_1 \le \sum_{k\in\mathbb{Z}} |\widehat{N_m}(\omega+k)|^2 \le C_2 \tag{3.10}$$

Proof. By construction $\widehat{N_m}(\omega) = (N_0(\omega))^m$. Moreover,

$$\widehat{N_0}(\omega) = \frac{e^{-2\pi i\omega}-1}{-2\pi i\omega} \quad (\omega \neq 0) \quad , \quad \widehat{N_0}(0) = 1 \quad .$$

In particular $\widehat{N_m}(\omega) \neq 0$ on $[-1/2, 1/2]$. Since the $\widehat{N_m}$ are continuous, we thus obtain $|\widehat{N_m}(\omega)|^2 \ge C_1 > 0$ on $[-1/2, 1/2]$, which entails the lower estimate of (3.10). For the other estimate we first observe that $(T_k N_0)_{k\in\mathbb{Z}}$ is in fact an ONS, which by 3.1.4 entails

$$\sum_{k\in\mathbb{Z}} |\widehat{N_0}(\omega+k)|^2 = 1 \quad ,$$

in particular $|\widehat{N_0}| \le 1$. Hence

$$\sum_{k\in\mathbb{Z}} \||\widehat{N_m}(\omega+k)|^2 = \sum_{k\in\mathbb{Z}} \||\widehat{N_0}(\omega+k)|^{2m} \le \sum_{k\in\mathbb{Z}} \||\widehat{N_0}(\omega+k)|^2 = 1 \quad ,$$

and the upper estimate is proved. □

The lemma now implies that the Box splines are in fact close to an ONB:

Lemma 3.4.5 $(T_k N_m)_{k\in\mathbb{Z}}$ *is a* **Riesz system**, *i.e., the mapping*

$$T : \ell^2(\mathbb{Z}) \ni (a_k)_{k\in\mathbb{Z}} \mapsto \sum_{k\in\mathbb{Z}} a_k T_k N_m$$

is a bijection between $\ell^2(\mathbb{Z})$ and $\widetilde{V_0} = \overline{\mathrm{span}(T_k N_m : k \in \mathbb{Z})}$, with

$$C_1 \|a\|_2^2 \le \|Ta\|_{\mathrm{L}^2}^2 \le C_2 \|a\|_2^2 \quad .$$

Here C_1 and C_2 are the constants from 3.4.4. Moreover, we have the characterization

$$f \in \widetilde{V_0} \Leftrightarrow \exists m_f \text{ 1-periodic with } \widehat{f} = m_f \widehat{N_m} \quad . \tag{3.11}$$

Proof. See [16, 2.8,2.10]. □

The following lemma shows how one computes from a Riesz system (as described in the previous lemma) an ONB, by a suitable orthonormalisation process.

Lemma 3.4.6 *Define*

$$\widehat{\varphi_m}(\omega) = \frac{\widehat{N_m}(\omega)}{\sqrt{\sum_{k\in\mathbb{Z}} |\widehat{N_m}(\omega+k)|^2}} \quad .$$

Then $(T_k \varphi_m)_{k\in\mathbb{Z}}$ is an ONB of $\widetilde{V_0}$.

Proof. (Sketch.) The construction of φ_m guarantees by 3.1.4 that $(T_k\varpi_m)_{k\in\mathbb{Z}}$ is an ONS. That it spans $\widetilde{V_0}$, follows from the fact that $\widehat{\varphi_m} = m_{\varphi_m} \cdot \widehat{N_m}$, where

$$m_{\varphi_m}(\omega) = \frac{1}{\sqrt{\sum_{k\in\mathbb{Z}} |\widehat{N_m}(\omega+k)|^2}}$$

is a 1-periodic function that is bounded from above and below. But this fact together with the characterisation (3.11) and Lemma 3.2.4 yields that $(T_k\varphi_m)_{k\in\mathbb{Z}}$ spans V_0. □

Proof of Theorem 3.4.2 (Sketch).

1. $\widetilde{V_0} = S^m(2^{-j}\mathbb{Z}) \cap \mathrm{L}^2(\mathbb{R})$ (see [16, 3.11])
2. Property (M6) is due to Lemma 3.4.6.
3. Property (M1) can be checked directly from the definitions.
4. The continuity of $\widehat{\varphi_m}$ and $\widehat{\varphi}(0) \neq 0$ are also checked by elementary arguments.

Hence Theorem 3.1.8 applies to show that $(V_j)_{j\in\mathbb{Z}}$ is an MRA. □

Chapter 4

Discrete Wavelet Transforms and Algorithms

This chapter contains the most important aspect of wavelets from the point of view of signal and image processing applications. The previous chapters have shown that wavelets are useful building blocks possessing a rather intuitive interpretation in terms of position and scale. What made wavelets popular was the realisation that there exist fast and simple algorithms which yield the decomposition into these building blocks. Moreover, the notion of multiresolution analysis, which at first appears like a clever but rather special construction, is the direct source of these algorithms.

For the efficient description of the algorithms we need some results concerning Fourier analysis on $\mathbb{Z}$.

4.1 Fourier transform on $\mathbb{Z}$

Definition 4.1.1 Given $a = (a_k)_{k\in\mathbb{Z}}$, we define the Fourier transform of a as a function on $[-1/2, 1/2]$ given by

$$\mathcal{F}(a)(\omega) = \sum_{k\in\mathbb{Z}} a_k e^{-2\pi i k\omega} \quad .$$

Here the sum converges in the L^2-sense; in fact, since the exponential functions are an ONB of $\mathrm{L}^2([-1/2, 1/2])$, the operator $\mathcal{F}$ is a unitary map $\ell^2(\mathbb{Z}) \to \mathrm{L}^2([-1/2, 1/2])$. We also use $\widehat{a} = \mathcal{F}(a)$. Depending on the context it may be convenient to regard $\mathcal{F}(a)$ as a periodic function on $\mathbb{R}$, or as living on a suitable interval of length 1. □

Definition 4.1.2 Given $f, g \in \ell^2(\mathbb{Z})$, we define their **convolution product** $f * g$ by

$$(f * g)(n) = \sum_{k\in\mathbb{Z}} f(k)g(n-k) \quad .$$

□

4.1.3 Exercises.

(a) The translation operator $T_k : \ell^2(\mathbb{Z}) \to \ell^2(\mathbb{Z})$ is given by $(T_k a)(n) = a(n-k)$. Show that $T_k f = f * \delta_k$, where $\delta_k(n) = \delta_{k,n}$.

(b) (Discrete differentiation.) A discrete analog to differentiation is given by the operator $(Df)(n) = \frac{f(n)-f(n-1)}{2}$. Compute $\widehat{Df}$.

(c) Let g be a sequence. Show that $f \mapsto f * g$ is a bounded operator on $\ell^2(\mathbb{Z})$ iff $g \in \ell^2(\mathbb{Z})$ with $\widehat{g} \in \mathrm{L}^\infty([-1/2, 1/2])$. Show that in this case the convolution theorem holds, i.e.,

$$(f * g)^\wedge = \widehat{f} \cdot \widehat{g} \quad .$$

(d) For $f \in \ell^2(\mathbb{Z})$ let f^* be defined as $f^*(k) = \overline{f(-k)}$. Compute $\widehat{f^*}$ and show that if the convolution operator $T : g \mapsto g * f$ is bounded, its adjoint is given by $T^* : g \mapsto g * f^*$.

□

Remark 4.1.4 It is obvious that on $\mathbb{Z}$ there are no arbitrarily large frequencies on $\mathbb{Z}$. In fact, it is not instantly clear whether one can define a meaningful ordering of frequencies from low to high: After all, depending on whether we regard the Fourier transform as living on $[-1/2, 1/2]$, different notions of "high" frequencies are the result. In the following, we want to make a point for taking $[-1/2, 1/2]$ as reference interval: The case $|\omega| = 1/2$ corresponds to the exponential $k \mapsto (-1)^k$. The oscillation between two neighboring points $k, k+1$ is 2, i.e., the maximal possible oscillation for exponentials. Conversely $\omega = 0$ corresponds to the constant function. These examples suggest that a meaningful comparison of frequencies is possible if we restrict to the interval $[-1/2, 1/2]$.

Another argument in favor of this view is obtained from the discrete differentiation operator. Recall from Section 1.6 that the continuous differentiation operator attenuates small frequencies and enhances large frequencies. Similarly, noting that $(Df) = f * d$, we obtain on the Fourier transform side $\widehat{Df} = \widehat{f} \cdot \widehat{d}$ for a suitable function d (which ?), and $|\widehat{d}|$ vanishes at zero and increases towards $\pm 1/2$. Hence D attenuates frequencies around zero, and enhances those close to $\pm 1/2$.

Finally, we mention that this point of view corresponds to a specific way of embedding $\ell^2(\mathbb{Z})$ into $\mathrm{L}^2(\mathbb{R})$, using the sampling theorem: The operator

$$S : (a_k)_{k \in \mathbb{Z}} \mapsto \sum_{k \in \mathbb{Z}} a_k T_k \mathrm{sinc}$$

is clearly isometric, with $(Sa)(k) = k$ and

$$\widehat{Sa}(\omega) = \begin{cases} \widehat{a}(\omega) & \omega \in [-1/2, 1/2] \\ 0 & \text{elsewhere} \end{cases}$$

□

4.2 The fast wavelet transform

Throughout this section we fix an MRA $(V_j)_{j \in \mathbb{Z}}$ with scaling function φ, wavelet ψ and scaling coefficients $(a_k)_{k \in \mathbb{Z}}$. Given $f \in V_0$, we can expand it with respect to two different ONBs:

$$\begin{aligned} f &= \sum_{k \in \mathbb{Z}} c_{0,k} T_k \varphi \\ &= \sum_{k \in \mathbb{Z}} d_{-1,k} D_2 T_k \psi + \sum_{k \in \mathbb{Z}} c_{-1,k} D_2 T_k \varphi \end{aligned}$$

Here we use the notations

$$\begin{array}{rcll} c_{j,k} & = & \langle f, D_{2^{-j}} T_k \varphi \rangle & \text{"approximation coefficient"} \\ d_{j,k} & = & \langle f, D_{2^{-j}} T_k \psi \rangle & \text{"detail coefficient"} \end{array}$$

In the following we will use $d^j = (d_{-j,k})_{k\in\mathbb{Z}}$ and $c^j = (c_{-j,k})_{k\in\mathbb{Z}}$. (Note the sign change in the index!) At the heart of the fast wavelet transform are explicit formulae for the correspondence

$$c^0 \leftrightarrow (c^1, d^1) \quad .$$

For this purpose we require one more piece of notation

Definition 4.2.1 (Up- and downsampling)
The up- and downsampling operators $\uparrow_2, \downarrow_2$ are given by

$$\begin{array}{rcll} (\downarrow_2 f)(n) & = & f(2n) & \text{("down")} \\ (\uparrow_2 f)(n) & = & \left\{ \begin{array}{cc} 0 & n \text{ odd} \\ f(k) & n = 2k \end{array} \right. & \text{("up")} \end{array}$$

We have $(\uparrow_2)^* = \downarrow_2$, as well as $\downarrow_2 \circ \uparrow_2 = \mathrm{Id}$. In particular, $\uparrow_2$ is an isometry. □

Theorem 4.2.2 *The bijection* $c^0 \leftrightarrow (c^1, d^1)$ *is given by*

$$\textit{Analysis:} \;\; c^{j+1} = \downarrow_2 (c^j * \ell) \quad , \quad d^{j+1} = \downarrow_2 (c^j * h)$$

and

$$\textit{Synthesis:} \;\; c^j = (\uparrow_2 c^{j+1}) * h^* + (\uparrow_2 d^{j+1}) * \ell^*$$

Here the filters h, ℓ *are given by* $\ell(k) = \frac{\overline{a_{-k}}}{\sqrt{2}}$ *and* $h(k) = \frac{(-1)^{1+k} a_{1+k}}{\sqrt{2}}$.

Proof. We first compute for $n, k \in \mathbb{Z}$ that

$$\begin{array}{rcl} \langle T_k \varphi, D_2 T_n \varphi \rangle & = & \langle T_{-n} D_2 T_k \varphi, \varphi \rangle \\ & = & \langle D_2 T_{k-2n} \varphi, \varphi \rangle \\ & = & \overline{\langle \varphi, D_2 T_{k-2n} \varphi \rangle} \\ & = & \dfrac{\overline{a_{k-2n}}}{\sqrt{2}} \quad . \end{array}$$

Hence

$$\begin{array}{rcl} c_{-1,n} = \langle f, D_2 T_n \varphi \rangle & & \\ & = & \displaystyle\sum_{k\in\mathbb{Z}} c_{0,k} \langle T_k \varphi, D_2 T_n \varphi \rangle \\ & = & \displaystyle\sum_{k\in\mathbb{Z}} c_{0,k} \frac{\overline{a_{k-2n}}}{\sqrt{2}} \\ & = & \displaystyle\sum_{k\in\mathbb{Z}} c_{0,k} \ell(2n-k) \\ & = & \left(c^0 * \ell\right)(2n) \quad . \end{array}$$

Analogous computations yield

$$\langle T_k\varphi, D_2T_n\psi\rangle = \frac{(-1)^{1-(k-2n)}a_{1-(k-2n)}}{\sqrt{2}}$$

and thus $d_{-1,n} = \left(c^0 * h\right)(2n)$.

The formula for synthesis is either checked directly, or by observing that the map $c^0 \mapsto (c^1, d^1)$ is unitary, hence inverted by its adjoint. The adjoint is easily computed using 4.1.3 and 4.2.1. □

Remarks 4.2.3 (a) **The "Cascade Algorithm:** Iteration of the analysis step yields c^j, d^j for $j \geq 1$. This can be sketched as the "cascade"

$$\begin{array}{ccccccccccc} c^0 & \rightarrow & c^1 & \rightarrow & c^2 & \rightarrow & \ldots & \rightarrow & c^{j-1} & \rightarrow & c^j \\ & \searrow & & \searrow & & \searrow & & \searrow & & \searrow & \\ & & d^1 & & d^2 & & & & d^{j-1} & & d^j \end{array}$$

whereas the synthesis step is given by

$$\begin{array}{ccccccccccc} c^j & \rightarrow & c^{j-1} & \rightarrow & c^{j-2} & \rightarrow & \ldots & \rightarrow & c^1 & \rightarrow & c^0 \\ & \nearrow & & \nearrow & & \nearrow & & \nearrow & & \nearrow & \\ d^j & & d^{j-1} & & & & d^2 & & d^1 & & \end{array}$$

The decomposition step corresponds to the orthogonal decompositions

$$\begin{array}{ccccccccccc} V_0 & \rightarrow & V_{-1} & \rightarrow & V_{-2} & \rightarrow & \ldots & \rightarrow & V_{-j+1} & \rightarrow & V_{-j} \\ & \searrow & & \searrow & & \searrow & & \searrow & & \searrow & \\ & & W_{-1} & & W_{-2} & & & & W_{-j+1} & & W_{-j} \end{array}$$

(b) **Complexity:** For a sequence $a = (a_k)_{k\in\mathbb{Z}}$ with finite support define the length of the support as

$$\mathcal{L}(a) = 1 + \max\{k : a_k \neq 0\} - \min\{k : a_k \neq 0\} \quad .$$

Now let a coefficient sequence c^0 be given with $\mathcal{L}(c^0) = N$, and assume that the filters h, ℓ are finite with length L. Then $\mathcal{L}(c^0 * h) = L + N - 1 = \mathcal{L}(c^0 * \ell)$, and thus

$$\mathcal{L}(c^1) = \mathcal{L}(\downarrow_2 (c^0 * \ell)) = \lfloor \frac{L + N - 1}{2} \rfloor \pm 1$$

and similarly for $\mathcal{L}(d^1)$. Thus (c^1, d^1) has only insignificantly more nonzero coefficients. (Note that "in practice" signals tend to be rather long, whereas wavelet filters usually are of the order of 10 to 30 coefficients.)

Moreover, the decomposition step $c^0 \mapsto (c^1, d^1)$ requires NL multiplications. Since in the iteration step only c^j is decomposed, we can estimate the complexity for the computation of $c^0 \mapsto (c^j, d^j, d^{j-1}, \ldots, d^1)$ by

$$\sum_{i=0}^{j-1} 2\mathcal{L}(c^i) \cdot L \leq 2\sum_{i=0}^{j-1}(2^{-i}N + L)L \quad .$$

Only $j \leq \log_2 N$ is meaningful, thus we finally see that the number of multiplications for the computation of $(c^j, d^j, d^{j-1}, \ldots, d^1)$ requires $O(NL)$ multiplications. The same holds for the inverse transform.

(c) **Interpretation:** By their origin, d^j and c^j can be interpreted as detail- resp. approximation coefficients of a certain function $f \in \mathrm{L}^2(\mathbb{R})$. However, the usual case in the applications of wavelets to signal processing is that the original data are given in form of the sequence c^0, to which the cascade algorithm is applied. The resulting coefficients can be interpreted in different ways:

(i) Associate the continuous-time function

$$f = \sum_{k \in \mathbb{Z}} c^0(k) T_k \varphi \quad ,$$

then the output of the cascade algorithm consists of detail and approximation coefficients. The problem with this approach is that φ is usually not known explicitly, and information about f is not easily accessible.

(ii) c^0 is considered as a sampled band-limited function: $c^0 = \widetilde{f}|_{\mathbb{Z}}$, $\mathrm{supp}\widehat{f} \subset [-1/2, 1/2]$, where

$$\widetilde{f}(x) = \sum_{k \in \mathbb{Z}} c^0(k) \mathrm{sinc}(x-k) \quad .$$

The projection of $\widetilde{f}$ onto V_0 can be computed via the scalar products $\widetilde{c}^0(k) = \langle \widetilde{f}, T_k \varphi,$ and feed $\widetilde{c}^0$ into the cascade algorithm. In short, we have recovered the interpretation of the output of the cascade algorithm by introducing a **preprocessing step** $c^0 \mapsto \widetilde{c}^0$.

In this context the **Coiflets**, introduced by Coifman, are particularly useful, since they are constructed such as to yield

$$\langle f, T_k \varphi \rangle \approx f(k)$$

for suitable functions f. In other words, using Coiflets one can regard c^0 as sampled values of a certain function f and at the same time interpret the output of the cascade algorithm as detail and approximation coefficients of f. (See [6] for details.)

(iii) Instead of referring to a suitable embedding into the continuous setting, one can also regard the transform $c^0 \mapsto (c^1, d^1)$ as an operator $\ell^2(\mathbb{Z}) \to \ell^2(\mathbb{Z}) \oplus \ell^2(\mathbb{Z})$, and study it as an operator of independent interest. The following discussion will be along this line. In particular, we will **not** assume that ℓ, h are obtained from a multiresolution analysis on $\mathbb{R}$. Instead, we investigate which pairs ℓ, h make the discrete wavelet transform work, and how the discrete wavelet transforms can be interpreted. Some of the arguments will look very similar to what we encountered in Chapter 3.

□

4.2.4 One-Step FWT as filterbank Let $(a_k)_{k \in \mathbb{Z}}$ be the scaling coefficients of an MRA and ℓ, h the filters associated to the MRA by 4.2.2. Denote by m_φ the mask associated to the scaling coefficients. Then a simple computation shows $\ell(\omega) = \sqrt{2 m_\varphi(-\omega)}$ and $\widehat{h}(\omega) = \sqrt{2 m_\psi(-\omega)}$. In particular, $m_\varphi(0) = 1$ and $m_\varphi(\pm 1/2) = 0$ imply that convolution with ℓ picks out the frequencies around 0 and attenuates frequencies around $\pm 1/2$; ℓ stands for "low frequencies". On the other hand, $\widehat{h}(0) = 0$ imply that convolution with h picks out the frequencies around $\pm 1/2$.

In summary, the pair (ℓ, h) acts as a filterbank on $\ell^2(\mathbb{Z})$. Again, this becomes most obvious for the Shannon MRA. □

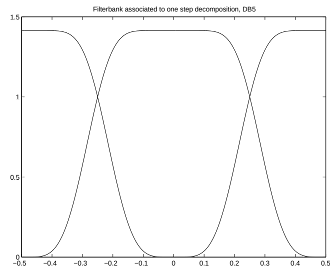

Figure 4.1: Plot of $|\widehat{\ell}|, |\widehat{h}|$ for the filters for the fifth Daubechies wavelet

Theorem 4.2.5 *Let $\ell, h \in \ell^2(\mathbb{Z})$ with bounded Fourier transforms. Let*

$$T : \ell^2(\mathbb{Z}) \to \ell^2(\mathbb{Z}) \times \ell^2(\mathbb{Z}) \quad , \quad c \mapsto (\downarrow_2 (c * \ell), \downarrow_2 (c * h)) \quad .$$

Then T is unitary iff

$$\text{(PR)} \qquad \left\{ \frac{1}{\sqrt{2}} \begin{pmatrix} \widehat{\ell}(\omega) \\ \widehat{\ell}(\omega + 1/2) \end{pmatrix}, \frac{1}{\sqrt{2}} \begin{pmatrix} \widehat{h}(\omega) \\ \widehat{h}(\omega + 1/2) \end{pmatrix} \right\} \subset \mathbb{C}^2 \text{ is an ONB}$$

holds for almost all $\omega \in [-1/2, 0]$. In this case, the inverse operator is given by

$$(c, d) \mapsto (\uparrow_2 c) * \ell^* + (\uparrow_2 d) * h^* \quad .$$

Note that property (PR) summarises the relations $(S3), (W1)$ and $(W1)$, as used in the proof of 3.2.5. (PR) is short-hand for "perfect reconstruction".

Proof. First note that

$$\begin{aligned} \downarrow_2 (c * \ell)(k) &= \sum_{m \in \mathbb{Z}} c(m) \ell(2k - m) \\ &= \sum_{m \in \mathbb{Z}} c(m) \overline{\ell^*(m - 2k)} \\ &= \langle c, T_{2k} \ell^* \rangle \quad , \end{aligned}$$

and by a similar computation $\downarrow_2 (c * h)(k) = \langle c, T_{2k} h^* \rangle$. Hence T is the coefficient map associated to the system

$$(+) \qquad (T_{2k} h^*)_{k \in \mathbb{Z}} \cup (T_{2k} \ell^*)_{k \in \mathbb{Z}} \quad ,$$

and we have to show that $(+)$ is an ONB. Now similar calculations as in the proof of 3.2.5 show that this is the case iff (PR) holds. □

Remark 4.2.6 Just as in Theorem 3.2.6 we can associate to any filter ℓ fulfilling the first half of (PR) in a canonical manner a filter h via

$$h(n) = (-1)^{1-n} \overline{\ell_{1-n}}$$

□

The following theorem can be found (with weaker assumptions) in [2].

Theorem 4.2.7 Wavelet-ONB in $\ell^2(\mathbb{Z})$
Let $\ell, h \in \ell^2(\mathbb{Z})$ be given with (PR). Assume in addition that ℓ is finitely supported. Given $c \in \ell^2(\mathbb{Z})$, define inductively

$$c^0 = c \quad , \quad c^{j+1} = \downarrow_2 (c^j * l) \quad , \quad d^{j+1} = \downarrow_2 (c^j * h) \quad .$$

(a) The **discrete wavelet transform**

$$W_d : c \mapsto (d^j(k))_{j \geq 1, k \in \mathbb{Z}}$$

is a unitary operator $\ell^2(\mathbb{Z}) \to \ell^2(\mathbb{N} \times \mathbb{Z})$.

(b) $d^j(k) = \langle c, T_{2^j k} \psi_j \rangle$ and $c^j(k) = \langle c, T_{2^j k} \varphi_j \rangle$, with suitable $\psi_j, \varphi_j \in \ell^2(\mathbb{Z})$ (for $j \geq 1$).

(c) ψ_j and φ_j can be computed recursively via

$$\begin{aligned} \varphi_0 &= \delta_0 \\ \varphi_{j+1} &= \varphi_j * \left(\uparrow_2^j \ell^* \right) \\ \psi_{j+1} &= \varphi_j * \left(\uparrow_2^j h^* \right) \end{aligned}$$

Proof. We first prove (b) and (c) by induction: Noting that $c^0 = c * \delta_0$, we find in the induction step

$$\begin{aligned} c^{j+1}(k) &= \downarrow_2 (c^* * \ell)(k) \\ &= \sum_{n \in \mathbb{Z}} c^j(n) \ell(2k - n) \\ &\overset{IH}{=} \sum_{n \in \mathbb{Z}} \langle c, T_{2^j n} \varphi_j \rangle \ell(2k - n) \\ &= \langle c, \sum_{n \in \mathbb{Z}} \overline{\ell(2k - n)} (T_{2^j n} \varphi_j) \rangle \quad . \end{aligned}$$

Now we can compute the

$$\begin{aligned} \sum_{n \in \mathbb{Z}} \overline{\ell(2k - n)} (T_{2^j n} \varphi_j)(m) &= \sum_{n \in \mathbb{Z}} \varphi_j(m - 2^j n) \overline{\ell(2k - n)} \\ &= \sum_{n \in \mathbb{Z}} \varphi_j(m - 2^j(n + 2k)) \overline{\ell(-n)} \\ &= \sum_{n \in \mathbb{Z}} \varphi_j(m - 2^j n - 2^{j+1} k)) \ell^*(n) \\ &= T_{2^{j+1} k} \varphi_{j+1} \end{aligned}$$

where φ_{j+1} is given as

$$\begin{aligned} \varphi_{j+1}(m) &= \sum_{n \in \mathbb{Z}} \varphi_j(m - 2^j n) \ell^*(n) \\ &= \sum_{n \in \mathbb{Z}} \varphi_j(m - n) (\uparrow_2^j \ell^*)(n) \\ &= (\varphi_j * (\uparrow_2^j \ell^*))(m) \quad . \end{aligned}$$

Replacing ℓ by h in the calculations yield the formula for ψ_j, and we have shown (b) and (c).

Now Theorem 4.2.5 implies that the mapping

$$c \mapsto (c^j, d^j, d^{j-1}, d^{j-2}, \dots, d^1)$$

is unitary, and hence by 1.2.10 the family

$$(T_{2^j k}\varphi_j)_{k\in\mathbb{Z}} \cup (T_{2^i k}\psi)_{1\le i\le j, k\in\mathbb{Z}}$$

is an ONB of $\ell^2(\mathbb{Z})$. Since this holds for all $j \geq 1$, we obtain in particular that $((T_{2^i k}\psi)_{1\le i\le j,k\in\mathbb{Z}}$ is an ONS in $\ell^2(\mathbb{Z})$. Hence the only missing property is totality. For this purpose define

$$P_j(f) := \sum_{k\in\mathbb{Z}} \langle f, T_{2^j k}\varphi_j\rangle T_{2^j k}\varphi_j \quad ,$$

which are the projections onto the orthogonal complement

$$((T_{2^i k}\psi)_{1\le i\le j,k\in\mathbb{Z}})^{\perp} \quad .$$

Hence we need to prove $P_j(f) \to 0$, for all $f \in \ell^2(\mathbb{Z})$. Since the space of finitely supported sequences is dense in $\ell^2(\mathbb{Z})$, it is enough to prove $P_j(f) \to 0$ for finite sequences (compare the proof of Theorem 3.1.8). For this purpose we need to auxiliary statements

- $\|\widehat{\varphi_j}\|_1 \to 0$, as $j \to \infty$. (Confer [2] for a proof.)
- $\mathcal{L}(\varphi_j) \leq \mathcal{L}(\ell) \cdot 2^j$. (This is an easy induction proof, using that $\mathcal{L}(\uparrow_2^j \ell^*) = 2^j \mathcal{L}(\ell) - 2^j$.)

As a result,

$$\begin{aligned} \|P_j(\delta_k)\|_2^2 &= \sum_{m\in\mathbb{Z}} |\langle \delta_k, T_{2^j m}\varphi_j\rangle|^2 \\ &= \sum_{m\in\mathbb{Z}} |\varphi_j(k - 2^j m)|^2 \\ &\leq (\mathcal{L}(\ell)+1)\,\|\varphi_j\|_\infty \\ &\leq (\mathcal{L}(\ell)+1)\,\|\widehat{\varphi_j}\|_1 \to 0 \quad , \end{aligned}$$

as $j \to \infty$. This concludes the proof of totality, hence (a) is shown. □

Remarks 4.2.8 (a) The family $(P_j)_{j\in\mathbb{Z}}$ of projections defines a decreasing sequence $V_j = P_j(\ell^2(\mathbb{Z}))$ of closed subspaces which share many properties of an MRA in $\mathrm{L}^2(\mathbb{Z})$. (Exercise: Check which properties are shared.)
(b) The finiteness of the filters are only used for totality of the system. Hence any pair ℓ, h with property (PR) yields an ONS in $\ell^2(\mathbb{Z})$.
(c) Note that the filterbank properties of the DWT, i.e., ℓ as low-pass and h as high-pass filter, enter nowhere in the proof (in fact, we could as well exchange the two). These are additional properties which we have to build into the filters.
(d) Theorem 4.2.7 implies that the map $c \mapsto d^j$ factors into a convolution with ψ_j^*, followed by a subsampling of 2^j. In particular, we can interpret the mapping

$$c \mapsto (d^1, d^2, d^3, \dots)$$

as a (subsampled) filterbank. Plotting $|\widehat{\psi_j}|^2$ allows to visualise how the different wavelets separates the different frequencies. For plotting purpose, it is useful to observe that on the Fourier transform side the recursion formulae read

$$\widehat{\varphi_{j+1}}(\omega) = \overline{\widehat{\ell}(2^j\omega)} \cdot \widehat{\varphi_j}(\omega) \quad , \quad \widehat{\psi_{j+1}}(\omega) = \overline{\widehat{h}(2^j\omega)} \cdot \widehat{\varphi_j}(\omega) \quad .$$

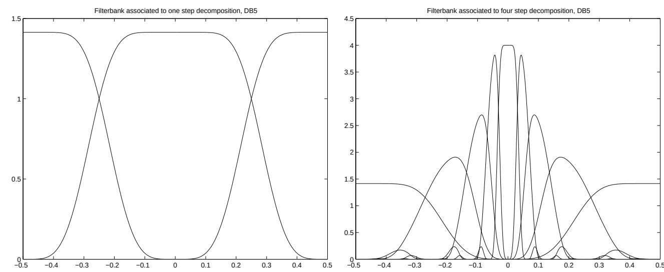

Figure 4.2: Filterbank plots for the fifth Daubechies wavelet and one and four iterations, respectively.

(e) Part (b) of the theorem shows that the fast wavelet transform computes the expansion coefficients of f with respect to an ONB of $\ell^2(\mathbb{Z})$. Once again, the elements of this ONB are indexed by scale and position parameters.

Note that the recursion formula in part (c) of the theorem is, up to indexing and normalisation, identical to the recursion formula obtained for the pointwise computation of the scaling function in Remark 3.3.11. Hence, whenever the filters ℓ, h are associated to an MRA in $\mathrm{L}^2(\mathbb{R})$, and in addition the recursion scheme in 3.3.11 converges (as it does for Daubechies wavelets), the large scale discrete wavelets $\psi_j \in \ell^2(\mathbb{Z})$ look more and more like the associated continuous wavelet ψ sampled on $2^{-j}\mathbb{Z}$. □

Proposition 4.2.9 *(Vanishing moments)*
Let ℓ, h be finite filters fulfilling the condition (PR). Assume that h is associated to ℓ as in remark 4.2.6. Then

$$\widehat{\ell}(\omega) = (e^{-2\pi i\omega} + 1)^r \widetilde{m}(\omega)$$

with $r \geq 1$ and $\widetilde{m}$ is a trigonometric polynomial. If $f \in \ell^2(\mathbb{Z})$ and $f|_I$ a polynomial of degree $< r$, then for all wavelets $T_{2^jk}\psi_j$ with support contained in I, we have

$$0 = \langle f, T_{2^jk}\psi_j \rangle = d^j(k) \quad .$$

Proof. The factorisation is obtained by the same argument as in the proof of 3.3.5. Hence $\widehat{h}(\omega) = \overline{\widehat{\ell}(\omega + 1/2)} e^{-2\pi i\omega}$ has a zero of order r at 0. Moreover, for finite filters ψ_j it is elementary to prove that

$$\sum_{k \in \mathbb{Z}} k^s \psi_j(k) e^{-2\pi i k\omega} = (2\pi i)^{-s} \widehat{\psi_j}^{(s)}(\omega) \quad .$$

But $\widehat{\psi_j}(\omega) = \widehat{h}(2^j\omega)\widehat{\varphi_{j-1}}(\omega)$ also has a zero of order r at 0, thus for $s < r$

$$0 = \frac{\widehat{\psi_j}^{(s)}(0)}{(2\pi i)^s} = \sum_{k \in \mathbb{Z}} k^s \psi_j(k)$$

which implies the statement of the proposition. □

4.3 2D Fast Wavelet Transform

The 2D Fast Wavelet Transform accounts for a large part of the popularity of wavelets in applications, specifically in image processing applications. In this section we outline the algorithm and its properties. The algorithm operates on elements of $\ell^2(\mathbb{Z}^2)$, which in the following are also called "(discrete) images". We will also speak of directions: The first coordinate is the horizontal coordinate, the second coordinate corresponds to the vertical direction.

Convolution and translation operators on $\ell^2(\mathbb{Z}^2)$ are defined in precisely the same way as in 1D.

The 1-step 2D fast wavelet transform can be defined as a concatenation of linewise and columnwise 1D wavelet decomposition steps. Assume for the following that we are given 1D discrete filters $\ell, h \in \ell^2(\mathbb{Z})$ which fulfill property (PR). For $f \in \ell^2(\mathbb{Z})$, we denote by $f(n_1, \cdot) \in \ell^2(\mathbb{Z})$ the restriction to $\{n_1\} \times \mathbb{Z}$. Define $f(\cdot, n_2)$ analagously. Then we define the one step decomposition of 2D FWT as an operator $S^2 : \ell^2(\mathbb{Z}^2) \to \ell^2(\mathbb{Z}^2)^4$, by the following operations:

- For $f \in \ell^2(\mathbb{Z}^2)$, define $c, d \in \ell^2(\mathbb{Z}^2)$ by

$$c(n_1, \cdot) = \downarrow_2 (f(n_1, \cdot) * \ell) \quad , \quad d(n_1, \cdot) = \downarrow_2 (f(n_1, \cdot) * h)$$

- define c^1 and $d^{1,j}$, $j = 1, 2, 3$ by

$$\begin{aligned} c^1(\cdot, n_2) &= \downarrow_2 (c(\cdot, n_2) * \ell) \\ d^{1,1}(\cdot, n_2) &= \downarrow_2 (c(\cdot, n_2) * h) \\ d^{1,2}(\cdot, n_2) &= \downarrow_2 (d(\cdot, n_2) * \ell) \\ d^{1,3}(\cdot, n_2) &= \downarrow_2 (d(\cdot, n_2) * h) \end{aligned}$$

- S^2 is defined by $S^2(f) = (c^1, d^{1,1}, d^{1,2}, d^{1,3})$.

We can rewrite the definition of S by use of the tensor product notation: If we define for $f, g \in \ell^2(\mathbb{Z})$ the tensor product $(f \otimes g)(n, k) = f(n)g(k)$, and in addition define the two-dimensional downsampling operator by

$$(\downarrow_2 f)(n, k) = f(2n, 2k)$$

then we may write

$$c^1 = \downarrow_2 (f * L) \quad , \quad d^{1,j} = \downarrow_2 (f * H_j)$$

where $L = \ell \otimes \ell$, $H_1 = \ell \otimes h$, $H_2 = h \otimes \ell$ and $H_3 = h \otimes h$. (Proof: Exercise.)

Similarly as in the case of the 1D wavelet transform, we obtain

Theorem 4.3.1 *Let $\ell, h \in \ell^2(\mathbb{Z})$ be finite filters satisfying property (PR). Then the operator S^2 is unitary. Hence the inverse operator is given by*

$$(c, d^1, d^2, d^3) \mapsto (\uparrow_2 c) * L^* + \sum_{i=1}^{3} \left(\uparrow_2 d^i\right) * H_i^* \quad .$$

Here $\uparrow_2$ is defined by

$$(\uparrow_2 f)(n,k) = \begin{cases} f(n/2,k/2) & (n/2,k/2) \in \mathbb{Z}^2 \\ 0 & \textit{otherwise} \end{cases}$$

Proof. Exercise. (Show, e.g., that S^2 factors into unitary maps, using the 1D case.) $\square$

4.3.2 Cascade Algorithm Again, the one-step decomposition can be iterated along the following scheme

$$\begin{array}{ccccccccccc} c^0 & \rightarrow & c^1 & \rightarrow & c^2 & \rightarrow & \ldots & \rightarrow & c^{j-1} & \rightarrow & c^j \\ & \searrow & & \searrow & & \searrow & & \searrow & & \searrow & \\ & & (d^{1,i})_{i=1,2,3} & & (d^{2,i})_{i=1,2,3} & & & & (d^{j-1,i})_{i=1,2,3} & & (d^{j,i})_{i=1,2,3} \end{array}$$

with associated inversion given by

$$\begin{array}{ccccccccccc} c^j & \rightarrow & c^{j-1} & \rightarrow & c^{j-2} & \rightarrow & \ldots & \rightarrow & c^1 & \rightarrow & c^0 \\ & \nearrow & & \nearrow & & \nearrow & & \nearrow & & \nearrow & \\ (d^{j,i})_{i=1,2,3} & & (d^{j-1,i})_{i=1,2,3} & & & & (d^{2,i})_{i=1,2,3} & & (d^{1,i})_{i=1,2,3} & & \end{array}$$

The column- and linewise definition of the wavelet transform allows a first interpretation of the resulting coefficients: Assuming that ℓ and h act as low- and highpass filters, we see that the different coefficient sets store different information contained in the image. More precisely:

- c^1 contains a low resolution copy of the image f, being obtained by low-pass filtering in both directions
- $d^{1,1}$ was obtained by low-pass filtering in the horizontal directions, followed by high-pass filtering in the vertical directions. As a result, only changes along the vertical direction are detected (e.g., edges in horizontal direction).
- $d^{1,2}$ records changes along the horizontal direction.
- $d^{1,3}$ records "diagonal" details.

Note that the properties concerning number of nonzero coefficients (for finite images) and algorithm complexity, which we observed in Remark 4.2.3 for the 1D case, remain intact in the 2D case. In fact, the implementation of the filters L, H_i is somewhat cheaper due to their tensor product structure: If ℓ, h have length $\mathcal{L}$, the associated 2D filters have size $\mathcal{L}^2$. On the other hand, first filtering horizontally with one filter and then vertically with the other one only requires $2\mathcal{L} \times$ (signal length) computation steps, not the expected $\mathcal{L}^2 \times$ (signal length). $\square$

The effect of iterating the one-step decomposition ad infinitum is described in the next theorem. It shows that the wavelet transform computes the expansion coefficients of the input signal with respect to a particular ONB of $\ell^2(\mathbb{Z})$.

Theorem 4.3.3 *Let $\ell, h \in \ell^2(\mathbb{Z})$ fulfill property (PR). Given $f \in \ell^2(\mathbb{Z}^2)$, let $d^{j,i}$,($j \in \mathbb{N}, i = 1,2,3$) denote the coefficients obtained by the cascade algorithm.*

(a) The **2D discrete wavelet transform**

$$W_d^2 : f \mapsto \left(d^{j,i}\right)_{j \in \mathbb{N}, i=1,2,3}$$

is a unitary map $\ell^2(\mathbb{Z}^2) \to \ell^2(\mathbb{N} \times \{1,2,3\} \times \mathbb{Z})$.

(b) Let $\psi_j, \varphi_j \in \ell^2(\mathbb{Z})$ *be the discrete-time wavelets and scaling functions computed in Theorem 4.2.7. Defining* $\Psi_{j,1} = \varphi_j \otimes \psi_j$, $\Psi_{j,2} = \psi_j \otimes \varphi_j$ *and* $\Psi_{j,3} = \psi_j \otimes \psi_j$. *Then*

$$d^{j,i}(k) = \langle f, T_{2^{-j}k} \Psi_{j,i} \rangle \quad .$$

Hence once again we obtain building blocks indexed by scale and position parameters, and a third parameter which can be interpreted as orientation.

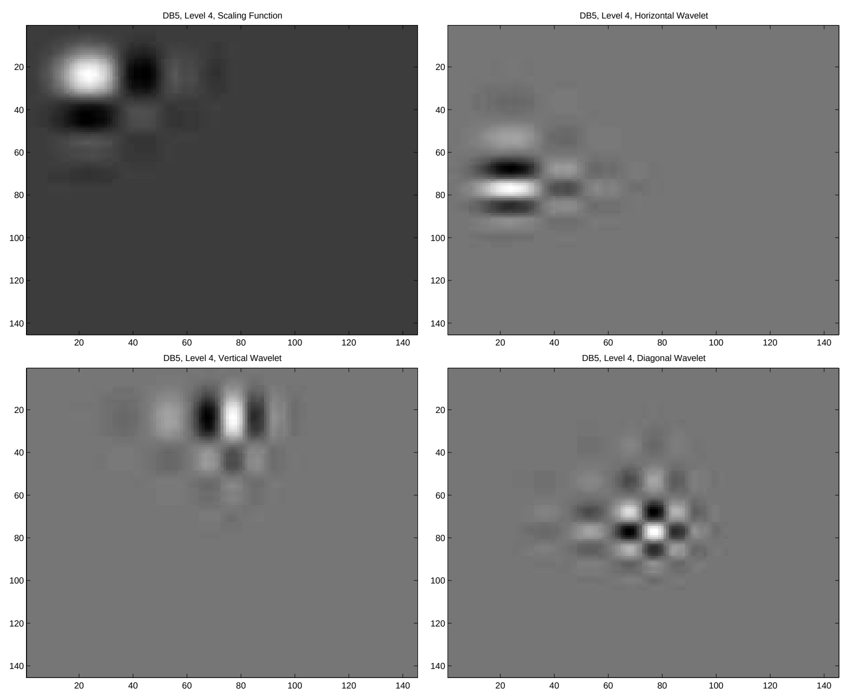

Figure 4.3: 2D scaling functions and wavelets at level 4, for the fifth Daubechies wavelet. From top left to bottom right: Scaling function, horizontal wavelet, vertical wavelet, diagonal wavelet

Remark 4.3.4 If $\ell, h \in \ell^2(\mathbb{Z})$ are associated to an MRA of $\mathrm{L}^2(\mathbb{R})$ with scaling function $\varphi \in \mathrm{L}^2(\mathbb{R})$ and wavelet $\psi \in \mathrm{L}^2(\mathbb{R})$, we can define a sequence of subspaces of $\mathrm{L}^2(\mathbb{R}^2)$ by letting

- $V_0 = \overline{\mathrm{span}\{T_k \varphi \otimes \varphi : k \in \mathbb{Z}^2\}}$
- $V_j = D_{2^{-j}} V_0$

It can be shown that these subspaces have all properties of an MRA (with $\mathbb{Z}^2$ replacing $\mathbb{Z}$). Also, the one-step fast wavelet transform can be associated to a change of ONB's in V_0, in just the same way that we derived the 1D FWT from an MRA of $\mathrm{L}^2(\mathbb{R})$ in the beginning of section 4.2. □

4.4 Signal processing with wavelets

In this section we sketch two standard applications of wavelets in signal processing, namely denoising and compression. For simplicity we restrict attention to the 1D case, though the discussion will show that the extension to 2D (e.g., images) is straightforward. Note that the description given here is just a very sketchy version of the real thing. For more details we refer to [6].

Roughly speaking, the use of wavelets for the two applications rests on two properties

- Existence of an inverse transform (orthogonality)
- "Meaningful" signals have only few big coefficients

The first property suggests a "transform domain processing" strategy: First compute the wavelet coefficients, do something clever to them, compute the inverse transform of the result. (Here the speed of transform and inverse transform algorithms is helpful, too.) The second property provides a heuristic for the (decisive) coefficient modification step. We will see below how the assumption can be exploited. Finally, orthogonality allows to control the ℓ^2 norm of the result.

Denoising deals with removal of (usually stochastic) noise from signals. Hence the following remark uses a number of very basic stochastic notions.

4.4.1 Denoising The denoising problem is described as follows: We are given a finite sequence x. x is called "observed signal"; we assume it to be of the form $x = y + n$. Here y is a (deterministic, unknown) signal, the "true signal", and n is a "noise term", i.e., the realisation of $\mathcal{L}(x)$ random variables. We assume that the $n(k)$ are independent, normally distributed with mean value zero and (common) variance σ^2. We do not assume that we know σ^2. Now the denoising problem may be formulated as: Given x, estimate y. In order to formalise this, we call any $x \to \Phi(x)$, where $\Phi(x)$ is a sequence with the same support as x, an estimator. We are looking for Φ which fulfills the **optimality criterion**

$$\text{Minimise the mean square error } \mathbb{E}(\|y - \Phi(x)\|^2)$$

As we already outlined above, we work in wavelet domain. More precisely, the proposed denoising algorithm consists of three steps:

1. Wavelet decomposition: $x \mapsto (c^j, d^j, \ldots, d^1)$.
2. $\widehat{d^i}(k) := \vartheta^i(k) d^i(k)$, for suitable factors $\vartheta^i(k)$
3. $(c^j, \widehat{d^j}, \ldots, \widehat{d^1}) \mapsto \widehat{x} =: \Phi(x)$

It turns out that it is possible to compute the optimal $\vartheta^i(k)$. (Note that this does **not** yield the minimal means square error estimator; only among those that fit into the above scheme.) In the following let $\varphi_{j,k} = T_{2^j k}\varphi_j$ and define $\psi_{j,k}$ analogously. Then we have

$$d^i(k) = \langle y, \psi_{j,k} \rangle + \langle n, \psi_{j,k} \rangle \quad ,$$

and thus we can compute

$$y - \Phi(x) = -\sum_{k \in \mathbb{Z}} \langle n, \varphi_{j,k} \rangle \varphi_{j,k} + \sum_{i,k} (1 - \vartheta^i(k)) \langle y, \psi_{i,k} \rangle \psi_{i,k} - \sum_{i,k} \theta^i(k) \langle n, \psi_{i,k} \rangle \psi_{i,k} \quad .$$

Hence, using linearity of the expected value and the unitarity of the wavelet transform we can compute

$$\mathbb{E}(\|y-\Phi(x)\|^2) = \sum_{k\in\mathbb{Z}} \mathbb{E}(|\langle n, \varphi_{j,k}\rangle|^2) + \sum_{k\in\mathbb{Z}, i=1,\ldots,j} \mathbb{E}\left(\left|(1-\vartheta^i(k))\langle y, \psi_{i,k}\rangle - \vartheta^i(k)\langle n, \psi_{i,k}\right|^2 \right) \quad .$$

This implies that we can optimise the $\vartheta^i(k)$ separately. In order to further simplify the term to be minimised, we make the following basic observations, which are due to vanishing moments of the wavelets and linearity of expected values:

(a) $\mathbb{E}(\langle n, \psi_{i,k}\rangle) = 0$ by linearity of the expected value, and since ψ_i has at least one vanishing moment.

(b) For arbitrary random variables y and constants $z \in \mathbb{C}$, $\mathbb{E}(|z+y|^2) = |z|^2 + 2\mathrm{Re}z\mathbb{E}(y) + \mathbb{E}(|y|^2)$.

(c) $\mathbb{E}(|\langle n, \psi_{i,k}\rangle|^2) = \sigma^2$.

Applying part (b) to

$$z = (1-\vartheta^i(k))\langle y, \psi_{i,k}\rangle \quad , \quad y = \vartheta^i(k)\langle n, \psi_{i,k}\rangle$$

we obtain that

$$\begin{aligned} &\mathbb{E}\left(\left|(1-\vartheta^i(k))\langle y, \psi_{i,k}\rangle - \vartheta^i(k)\langle n, \psi_{i,k}\right|^2 \right) = \\ &= |(1-\vartheta^i(k))\langle y, \psi_{i,k}\rangle|^2 + |\vartheta^i(k)|^2 \mathbb{E}(|\langle n, \psi_{i,k}\rangle|^2) \\ &= |(1-\vartheta^i(k))|^2 |\langle y, \psi_{i,k}\rangle|^2 + |\vartheta^i(k)|^2\sigma^2 \quad . \end{aligned}$$

Recall that we need to minimise this as a function of $\vartheta^i(k)$. But now this is just basic calculus (computing zeroes of the derivative) that yields the optimal value

$$\vartheta^i(k) = \frac{|\langle y, \psi_{i,k}\rangle|^2}{|\langle y, \psi_{i,k}\rangle|^2 + \sigma^2} \quad .$$

Hence we have derived that the optimal weights are between 0 and 1. Note however that we have expressed $\vartheta^i(k)$ in terms of two unknown quantities, σ^2 and y.

For the estimation of σ^2 we use an assumption on y: $\langle y, \psi_{i,k}\rangle = 0$ for most k. (This holds for instance if y is piecewise polynomial and ψ_1 has a sufficient number of vanishing moments.) Then

$$|\langle x, \psi_{i,k}\rangle| = |\langle n, \psi_{i,k}\rangle| \quad ,$$

for most k. We know that the random variables $\langle n, \psi_{i,k}\rangle$ are mean zero, independent and normally distributed with variance σ^2. An estimator taken from the literature is

$$\widehat{\sigma} = \frac{\mathrm{Med}(|\langle x, \psi_{i,k}\rangle|)}{0.6745} \quad .$$

Here $\mathrm{Med}(x_1, \ldots, x_n)$ is the **median** defined as $y \in \mathbb{R}$ such that $y \geq x_i$ for $n/2$ indices and $y \leq x_i$ for $n/2$ indices. (Simply sort the x_i in ascending order and take for y the middle element.) The choice of the median is motivated by the following: It is more robust with regard to outliers, which in our setting would correspond to those k for which $\langle y, \psi_{1,k}\rangle \neq 0$.

For the computation of the $\vartheta^i(k)$ we would still have to know the scalar products $\langle y, \psi_{i,k}\rangle$. We replace this optimal choice by a simple thresholding scheme, namely fix a threshold

$$T = \sigma\sqrt{2\ln N}$$

and let

$$\vartheta^i(k) = \begin{cases} 1 & |\langle x, \psi_{1,k}| \geq T \\ 0 & \text{otherwise} \end{cases}$$

This thresholding scheme was proposed by Donoho and Johnston, who proved error estimates and showed that asymptotically the algorithm yields optimal results.

We have thus obtained the following **wavelet thresholding algorithm**:

1. Compute the wavelet coefficients, $x \mapsto (c^j, d^j, d^{j-1}, \ldots, d^1)$.
2. Compute the threshold

$$T_i = \frac{\text{Med}(|d^i(k)| : k \in \mathbb{Z})\sqrt{2\ln N}}{0.6745} \quad .$$

3. Define

$$\widehat{d^i}(k) = \begin{cases} d^i(k) & \text{if } d^i(k) > T_i \\ 0 & \text{otherwise} \end{cases}$$

4. Compute wavelet inversion: $(c^j, \widehat{d^j}, \ldots, \widehat{d^1}) \mapsto \Phi(x)$.

□

4.4.2 Compression Lossy compression is a procedure which replaces an original signal (byte sequence, grey level image, etc.) by an approximation which is simpler to store (i.e., needs less bits).

The naive approach to wavelet based compression is this: For any "decent" (e.g., piecewise polynomial) signal x and most wavelets $\psi_{j,k}$ we have $\langle x, \psi_{j,k}\rangle \approx 0$. Hence a simple compression scheme would consist of the following steps:

- Pick a compression rate $r > 1$
- Store only the N/r biggest wavelet coefficients (N = signal length)
- The signal reconstructed from these coefficients yields an approximation of x which has the smallest ℓ^2 error among all approximations with N/r terms.

While this proposal captures one idea behind wavelet based compression, it is unrealistic since it ignores the fact that real numbers can only be stored approximately, and that the right choice of assigning bits to this task is essential to compression.

In addition, it is not enough to store the values of the coefficients but also their position/index. Note that for an image of size 256×256, the indexing takes 16 bits, whereas single pixel values are often only coded in 8 bits. This shows that indexing may not be neglected.

A more elegant solution, which is based on coding theoretic notions, is the following lossy compression scheme, which we will explain in the following.

1. Compute the wavelet coefficients, $x \mapsto (c^j, d^j, d^{j-1}, \ldots, d^1)$.

2. $(c^j, d^j, \ldots, d^1) \mapsto (\widetilde{c^j}, \widetilde{d^j}, \ldots, \widetilde{d^1})$, where the $\widetilde{c^j}$ and $\widetilde{d^i}$ are quantised (rounded) values; typically

$$\widetilde{d^j}(k) = \frac{m^j(k)}{\Delta_j} \approx d^j(k)$$

for some integer numbers $m^j(k)$, and stepsizes Δ_j.

3. Code the values Δ_j and the integers used to approximate d^i and c^j (e.g., by a Huffman code), store the result.

The compression rate is controlled by the stepsizes Δ_j; the larger Δ_j, the smaller the number of symbols used to approximate wavelet and scaling coefficients. Computing the image (or rather: its compressed approximation $\widetilde{x}$) from the stored data is a simple two-step procedure:

1. Read and decode the approximated coefficients $(\widetilde{c^j}, \widetilde{d^j}, \ldots, \widetilde{d^1})$

2. Compute the inverse wavelet transform $(\widetilde{c^j}, \widetilde{d^j}, \ldots, \widetilde{d^1}) \mapsto \widetilde{x}$

To understand how this scheme compresses data, we need a few notions from coding theory, which will now be explained (rather sketchily). Suppose that we are given an "alphabet" $\mathcal{A}$, which is a finite set, and a "word" $W = a_1 \ldots a_N \in \mathcal{A}^N$. A mapping $\Phi : a \mapsto \Phi(b)$, assigning to each element of the alphabet a sequence of bits (possibly of varying length), is called "coding". (We ignore questions of uniqueness, prefix-codes etc. for reasons of simplicity.) We are looking for a coding for which the bit-sequence $\widetilde{W} = \Phi(b_1)\Phi(b_2)\ldots\Phi(b_N)$ has minimal length.

There exist theoretical bounds for the length of $\widetilde{W}$. For $a \in \mathcal{A}$ let

$$p(a) = \frac{\sharp\{i : a_i = a\}}{N}$$

denote its relative frequency in the word W. We define the **entropy** of W as

$$H(W) = -\sum_{a \in \mathcal{A}} p(a) \log(p(a))$$

using the convention $0 \log 0 = 1$. Then there is the following

Theorem: (Shannon) The optimal coding yields a length $\widetilde{W} \approx N\left(\frac{H(W)}{\log 2} + const\right)$, for a (small) constant *const*.

The key idea behind the theorem is this: The number of bits used to store elements of the alphabet should depend on the relative frequency. Those that occur very often should use less bits. If all symbols occur with the same frequency, nothing can be gained. Note that the maximiser of the entropy functional is given by the uniform distribution, i.e., it is attained whenever all elements of the alphabet occur with equal frequency, with associated entropy $\log(N)$. Note also that the stupid way of storing, i.e., every symbol gets the same number of bits, results in a length of $N \log_2 |\mathcal{A}|$, and thus $H(W)$ can be interpreted as a compression rate.

Now we have the means to quantify the tradeoff governing wavelet compression: Computing and quantising the coefficients results in a loss of information, which can be measured by the ℓ^2-error incurred by quantising (note that here the ONB property of wavelets enters). On the other hand, the transform step results in a decrease of entropy, which by Shannon's theorem can be directly translated into a compression rate. □

Bibliography

[1] C. Blatter, *Wavelets. Eine Einfhrung.* Vieweg, 1998

[2] A. Cohen and R.D. Ryan, *Wavelets and Multiscale Signal Processing.* Chapman & Hall, 1995.

[3] I. Daubechies, Ten Lectures on Wavelets.

[4] M. Holschneider, *Wavelets. An Analysis Tool.* Clarendon Press, 1995.

[5] A. Louis, P. Maass and A. Rieder, *Wavelets. Theorie und Anwendungen.* Teubner Studienbücher: Mathematik. 1998.

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

[7] S. Mallat and W.-L. Hwang,*Singularity detection and processing with wavelets.* IEEE Trans. Inf. Theory **38**, 617-643 (1992).

[8] D. Marr, *Vision.* Freeman (New York) 1982.

[9] Y. Meyer, *Wavelets and Operators.* Cambridge University Press 1995.

[10] R. Coifman and Y. Meyer, *Wavelets. Calderón-Zygmund and multilinear operators.* Cambridge University Press 1997

[11] Y. Meyer, *Wavelets. Algorithms and Applications.* SIAM, 1993.

[12] O. Rioul, *Simple regularity criteria for subdivision schemes.* SIAM J. Math. Anal. **23**, 1544-1576 (1992).

[13] W. Rudin, *Real and Complex Analysis.3rd ed.* McGraw Hill, 1987.

[14] B. Torrésani, *Analyse continue par ondelettes* . InterÉditions Paris, 1995.

[15] R. Wheeden and A. Zygmund, *Measure and Integral. An introduction to real analysis.* Marcel Dekker, 1977.

[16] P. Wojtaszczyk, *A Mathematical Introduction to Wavelets.* London Mathematical Society Student Texts 37. Cambridge University Press, 1997.