Harmonic Analysis: from Fourier to Haar
Lecture Notes delivered at the Program for Women in Mathematics
Institute for Advanced Study, Princeton
May 17-28, 2004

María Cristina Pereyra* Lesley Ward †

June 9, 2004

Contents

*Department of Mathematics and Statistics, University of New Mexico, Albuquerque, NM 87131, USA; `http://www.math.unm.edu/~crisp`, `crisp@math.unm.edu`.

†Department of Mathematics, Harvey-Mudd College, Claremont, CA 91711, USA; `http://www.math.hmc.edu/~ward`, `ward@math.hmc.edu`.

1 Introduction

This two week course will try to expose the students to the basics of harmonic analysis: from Fourier's heat equation, and the decomposition of functions into sums of cosines and sines (frequency analysis) to the dyadic harmonic analysis (or decomposition into Haar basis, space localization). In between these two different ways of decomposing functions there is a whole world of time/frequency analysis (wavelets) which we will touch on, although we will concentrate on the Fourier and Haar cases.

The first week, some of the classical results on Fourier series were discussed thoroughly including different modes of convergence, Gibbs phenomenon, interplay with differentiability, as well as the conceptually simpler case of Fourier analysis on finite dimensional spaces and the celebrated Fast Fourier Transform, an algorithm discovered by Gauss almost 200 years ago and rediscovered by Cooley and Tuckey in the 60s.

The second week, we will discuss the Fourier transform on the line, as well as survey the windowed Fourier transform and the wavelet transform. The Haar basis, the geometry of dyadic intervals, and the multiresolution analysis that they induce will be analysed carefully. We will describe some design features and basic properties of known wavelets, as well as the very basics of image/signal denoising and compression. The Hilbert transform, probably the most important operator in harmonic analysis after the Fourier transform, as well as its connections to complex and Fourier analysis will be introduced. We will discuss the behavior of the Hilbert transform in the function space L^p, as well as some tools to understand these spaces. In particular we will look at some basic interpolation results, and as a consequence we will derive some of the most useful inequalities in analysis. Time permitting, we will show how to express the Hilbert transform as a superposition of dyadic operators, and we will analyse the decay of its coefficients in terms of the Haar basis.

The problem sessions will be oriented toward applications and sometimes careful calculations avoided in class. We will use some of the software available like the Matlab Wavelets Toolbox, to illustrate the power of these techniques in the "real world".

Prerequisites: advanced calculus and linear algebra. We will introduce as needed concepts from Hilbert spaces, Banach spaces, distributions. In particular we will try to compare between finite and infinite dimensional spaces and remark on their differences and their analogies.

We would like to thank Sun-Yung Alice Chang and Karen Uhlenbeck, the organizers of the 2004 Program for Women in Mathematics on *Analysis and Partial Differential Equations*, for inviting us to deliver these lectures. The program is now in its 11th edition, and has been very successful drawing bright young women (undergraduate, graduate, and postdoctorates) into a two week intensive session in Mathematics at the Institute for Advanced Study in Princeton.

Last but not least, we would like to thank our TAs, Manuela de Castro and Stephanie Molnar for their invaluable help, and all the students for their enthusiasm and vitality.

The pictures in these notes were kindly provided by Martin Mohlenkamp.

2 Crash course on the Fourier transform on the line

The key idea of harmonic analysis is to represent signals as a superposition of simpler functions that are well understood. Traditional Fourier series represent periodic functions as a sum of pure harmonics (sines and cosines),

$$f(x) = \sum_{n\in\mathbb{Z}} \hat{f}(n) e^{2\pi i n x}. \tag{1}$$

Fourier's statement in the early 1800's that *any periodic function* could be expanded in such a series, where the coefficients (amplitudes) for each frequency n are calculated by

$$\hat{f}(n) = \int_0^1 f(x) e^{-2\pi i n x} dx, \tag{2}$$

revolutionized mathematics. It took almost 150 years to settle on exactly what this meant. Only in 1966, Lenhart Carleson in a remarkable paper [Car], showed that for square integrable functions on $[0,1]$ the Fourier partial sums converge pointwise a.e., and as a consequence the same holds for continuous functions (this was unknown until then).

You have discussed these issues and more in the past week. Among other things, you showed that:

The exponential functions $\{e^{2\pi i n x}\}_{n\in\mathbb{Z}}$ *form an orthonormal basis in* $L^2([0,1])$.

In the non-periodic setting one has a *Fourier* and an *inverse Fourier* integral transforms

$$\hat{f}(\xi) = \int_{\mathbb{R}} f(x) e^{-2\pi i \xi x} dx, \quad \check{f}(\xi) = \int_{\mathbb{R}} f(x) e^{-2\pi i \xi x} dx$$

There is much to be said about for which functions these formulae make sense, and for which of them, and in what sense, the following *Fourier inversion formula* holds, $\check{\hat{f}} = f$, i.e.,

$$f(t) = \int_{\mathbb{R}} \hat{f}(\xi) e^{2\pi i \xi t} d\xi.$$

We will digress over these issues in today's lecture.

2.1 From Fourier series to Fourier integrals

Heuristically we could arrive to the integral formulae above by calculating the Fourier series on larger and larger intervals, until we "cover the whole line". This was Fourier's approach: first calculate the Fourier series of a "nice" function f, on the symmetric interval $(-L/2, L/2)$,

$$f(x) = \sum_{n\in\mathbb{Z}} a_n(L) e^{2\pi i n x/L}, \quad t \in (-L/2, L/2), \tag{3}$$

where the coefficients are given by the formula,

$$a_n(L) = \frac{1}{L}\int_{-L/2}^{L/2} f(x)e^{-2\pi inx/L}dx.$$

Then, let

$$\xi_n = n/L, \quad \text{so that} \ \ \Delta\xi = \xi_{n+1} - \xi_n = 1/L,$$

and consider the function

$$F_L(\xi) = \int_{-L/2}^{L/2} f(x)e^{-2\pi i\xi x}dx.$$

We can now rewrite the Fourier series in (3) as,

$$f(x) = \sum_{n\in\mathbb{Z}} F_L(\xi_n)e^{2\pi i\xi_n x}\Delta\xi.$$

This could be viewed as a formal Riemann sum for the integral $\int_{-L/2}^{L/2} F_L(\xi)d\xi$, except that the parameter L appears in the function to be integrated, in the limits of integration, as well as in the partition of the interval. Let us be careless about that. Let $\Delta\xi \to 0$ (that is, $L \to \infty$), and after observing that $\lim_{L\to\infty} F_L(\xi) = \hat{f}(\xi)$, we could perhaps accept that we need *all* frequencies to recover a function in $\mathbb{R}$, and that the following integrals will play the same role as (1), and (2) do in the Fourier series theory,

$$f(x) = \int_{\mathbb{R}} \hat{f}(\xi)e^{2\pi i\xi x}d\xi, \quad \hat{f}(\xi) = \int_{\mathbb{R}} f(x)e^{-2\pi i\xi x}dx.$$

To make this argument rigorous, we have to specify what "nice" means, so that we have pointwise convergence, or some other type of convergence, and we are entitled to perform all these manipulations. See for example, Exercise 1 in [StSh, Ch. 5].

2.2 Fourier transform in $\mathcal{S}$

It turns out that these formulae hold when f belongs to the *Schwartz space*, $\mathcal{S}(\mathbb{R})$. This space consists of infinitely differentiable functions $f : \mathbb{R} \to \mathbb{C}$ (denoted $C^\infty(\mathbb{R})$), so that f and all its derivatives f', f'', ..., $f^{(\ell)}$,..., are *rapidly decreasing.* That is, $\mathcal{S}(\mathbb{R})$ is made up of those functions $f \in C^\infty(\mathbb{R})$ for which,

$$\sup_{x\in\mathbb{R}} |x|^k|f^{(\ell)}(x)| < \infty \quad \text{for all integers} \ \ k, \ell \geq 0.$$

The Schwartz space is a vector space over the complex numbers, that is closed under multiplication by polynomials and differentiation. Functions in $\mathcal{S}(\mathbb{R})$ decay at $\pm\infty$ faster than any polynomial, in particular, compactly supported $C^\infty(\mathbb{R})$ functions are in $\mathcal{S}(\mathbb{R})$. In the Schwartz space, we have a suitable integration theory, just by considering the integral over $\mathbb{R}$ to be the limit of Riemann integrals over larger and larger intervals. It does make sense to talk about Fourier transforms of Schwartz functions. For $f \in \mathcal{S}(\mathbb{R})$, define its Fourier transform $\hat{f} : \mathbb{R} \to \mathbb{C}$, by the integral formula advertised at the beginning of the lecture,

$$\hat{f}(\xi) = \int_{\mathbb{R}} f(x) e^{-2\pi i \xi x} dx = \lim_{N \to \infty} \int_{-N}^{N} f(x) e^{-2\pi i \xi x} dx.$$

Example 1 *The canonical example of a function in $\mathcal{S}(\mathbb{R})$ is the* Gaussian *defined by*

$$K(x) = e^{-\pi x^2}.$$

The Gaussian plays a very important role in Fourier analysis, probability theory and physics. It has the property that it equals its Fourier transform,

$$\hat{K}(\xi) = e^{-\pi \xi^2}.$$

This can be proved using the observation that both the Gaussian and its Fourier transform satisfy the ordinary differential equation $f'(x) = -2\pi x f(x)$, with initial condition $f(0) = 1$, hence by uniqueness of the solution they must be the same function!

Exercise 1 *Convince yourself that the Gaussian belongs to $\mathcal{S}(\mathbb{R})$. Find the Fourier transform of the Gaussian, by either filling in the details in the previous paragraph, or by a direct calculation. (The time/frequency dictionary in Section* 2.3 *might come handy, in particular property* (h)*.)*

Exercise 2 *Find a Schwartz function of compact support* (a bump function)*. See Exercise* 4 *in* [StSh, Ch. 5].

2.2.1 Convolutions

A very important operation in harmonic analysis is the *convolution*, as you have already discussed in the context of Fourier series. On the line, and for functions $f, g \in \mathcal{S}(\mathbb{R})$, it is defined as follows,

$$f * g(x) = \int_{\mathbb{R}} f(x-y) g(y) dy.$$

The integral defines a rapidly decreasing function, due to the smoothness and fast decay of the functions f and g. It turns out that the convolution is a closed operation in $\mathcal{S}(\mathbb{R})$, and the key observation to believe this fact is that

$$\frac{d}{dx}\big(f * g\big)(x) = f' * g(x) = f * g'(x).$$

The first identity above is true because in this setting, we are entitled to interchange differentiation and integration. The second identity is true by integration by parts, or by noting that the convolution is a *commutative* operation.

Convolution is a *smoothing* operation. The output keeps the best of each input. This heuristic is not completely clear when thinking about convolution of Schwartz functions, because you are already in the best world. One can convolve much less regular functions (e.g. just an integrable function) with a smooth function, then the smoothness will be transferred to the convolution.

Convolutions correspond to *translation invariant bounded linear transformations*, and to *Fourier multipliers*. These, and their generalizations, have been and are the object of intense study.

2.2.2 Approximations of the identity

Convolution can be viewed as a group operation, as it is commutative, but one can check that there is no *unit* in the group (the unit would be the delta function, to be defined in a couple of sections, which is not a true function but a distribution or, if you prefer, a point-mass). A substitute for the lack of unit would be the an *approximation of the identity*, which you have already discussed in the context of Fourier series.

An *approximation of the identity* in $\mathbb{R}$ is a family of functions $\{k_t(x)\}$, t in some directed index set, such that,

$$\text{(i)} \quad \int_{\mathbb{R}} k_t(x)dx = 1,$$

$$\text{(ii)} \quad \int_{\mathbb{R}} |k_t(x)|dx < M, \text{ for all } t.$$

$$\text{(iii)} \quad \lim_t \int_{|x|>\eta} |k_t(x)|dx = 0.$$

In the examples it will become clear how the limit is taken. Let t_0 be an accumulation point of the index set, then we consider $t \to t_0$. Sometimes $t > 0$ then $t \to 0$, other times, $t < \infty$ then $t \to \infty$.

Example 2 *Gaussians are good kernels, and by scaling we obtain an* approximation of the identity. *More precisely, define for* $K(x) = e^{-\pi x^2}$, *the Gauss kernel,* $K_t(x) = t^{-1}K(xt^{-1})$, *then this family satisfies the above properties when* $t \to 0$. *The graphs of* $K_t(x)$ *look more and more like the delta-function as* $t \to 0$.

Convolving against an approximation of the identity one can approximate a function very well. One such instance of this heuristic is given in the next exercise, whose analogue on the circle you already discussed.

Exercise 3 *If* $f, K \in \mathcal{S}(\mathbb{R})$, *and* $\{K_\delta(x) = \delta^{-1}K(x\delta^{-1})\}_{\delta>0}$ *is an approximation of the identity (i.e.* (i)-(iii) *are satisfied), then the convolution* $f * K_\delta(x)$ *converges to* $f(x)$ *uniformly in* x *as* $\delta \to 0$. *See Corollary* 1.7 *in* [StSh, Ch. 5].

2.2.3 Fourier Inversion Formula and Plancherel

The Fourier transform interacts very nicely with a number of operations. In particular differentiation is transformed into polynomial multiplication and viceversa, which "explains" the immense success of the Fourier transform techniques in the study of differential equations (certainly in the linear case). Convolutions are transformed into products which explains the success in signal processing, as "filtering" (=convolution) is one of the most important signal processing tools. In Section 2.3, you can find a time/frequency dictionary which lists all these useful interactions.

As a consequence of differentiation being transformed into polynomial multiplication, and viceversa, we can check that *the Fourier transform maps $\mathcal{S}(\mathbb{R})$ into itself.*

All of these facts, together with the *multiplication formula*[1]

$$\int_{\mathbb{R}} f(x)\hat{g}(x)dx = \int_{\mathbb{R}} \hat{f}(y)g(y)dy, \tag{4}$$

which are valid for functions $f, g \in \mathcal{S}(\mathbb{R})$, can be weaved together to provide a proof of the Fourier inversion formula in $\mathcal{S}(\mathbb{R})$.

Theorem 1 (Fourier Inversion formula in $\mathcal{S}$) *If $f \in \mathcal{S}(\mathbb{R})$, then for all $t \in \mathbb{R}$,*

$$f(t) = \int_{\mathbb{R}} \hat{f}(y)e^{2\pi iyt}dy.$$

Proof: Apply (4) with

$$g(x) = e^{2\pi ixt}e^{-\pi|\delta x|^2}.$$

By properties (c) and (d) in the time/frequency dictionary in Section 2.3,

$$\hat{g}(x) = \delta^{-1}e^{-\pi|(t-x)/\delta|^2} = K_\delta(t-x), \tag{5}$$

which is the approximation of the identity generated by the Gaussian. Now (4) reads,

$$\int_{\mathbb{R}} f(x)K_\delta(t-x)dx = \int_{\mathbb{R}} \hat{f}(y)e^{2\pi iyt}e^{-\pi|\delta y|^2}dy.$$

Let $\delta \to 0$. The left-hand-side converges to $f(t)$ uniformly by Exercise 3. As for the right-hand-side, we are in a setting where we are entitled to interchange limit and integral, hence as $\delta \to 0$, it converges to $(\hat{f})^\vee(t)$, which is the inversion formula we were seeking. See [StSh, Ch 5, Sec. 1.5] for more details. ◇

The inversion formula guarantees that the map is in fact a *bijection* on the Schwartz space. More can be said: it is an *energy preserving map* or an *unitary transformation* on $\mathcal{S}(\mathbb{R})$. We can equip $\mathcal{S}$ with an inner product

$$\langle f, g\rangle = \int_{\mathbb{R}} f(x)\overline{g(x)}dx$$

[1] This is a consequence of the Fubini-Tonelli Theorem for interchange of integrals.

with associated norm

$$\|f\|_2 = \left(\int_{\mathbb{R}} |f(x)|^2 dx\right)^{1/2}.$$

Theorem 2 (Plancherel) *If $f \in \mathcal{S}(\mathbb{R})$ then $\|f\|_2 = \|\hat{f}\|_2$.*

Exercise 4 *Prove Plancherel's identity for Schwartz functions with compact support, following the heuristic ideas at the beginning of the lecture (that is use Plancherel's identity for Fourier series on larger and larger intervals and take limits). Prove the* polarization identity *for real-valued $f, g \in \mathcal{S}(\mathbb{R})$, namely,*

$$\langle f, g\rangle = \langle \hat{f}, \hat{g}\rangle. \tag{6}$$

Hint: *Use Plancherel for $f+g$ and $f-g$ then add.*

2.2.4 Kernels' Interlude

We used the Abel and Césaro means to regularize the convergence of Fourier series, by convolving with the Poisson and Fejér kernels on the circle, which happened to be approximations of the identity on the interval $[0,1]$. We can define analogous kernels on $\mathbb{R}$, which can be used to regularize the convergence of the Fourier integral.

Fejér kernel on $\mathbb{R}$:

$$F_R(x) = R\left(\frac{\sin(\pi R x)}{R\pi x}\right)^2, \quad 0 < R < \infty. \tag{7}$$

Poisson kernel on $\mathbb{R}$:

$$P_y(x) = \frac{y}{\pi(x^2+y^2)}, \quad y > 0. \tag{8}$$

The Poisson kernel is a solution of *Laplace's equation*, $\Delta P = \dfrac{\partial^2 P}{\partial x^2} + \dfrac{\partial^2 P}{\partial y^2} = 0.$

Heat kernel on $\mathbb{R}$ (a modified Gaussian):

$$H_t(x) = \frac{1}{\sqrt{4\pi t}} e^{-\frac{|x|^2}{4t}}, \quad t > 0. \tag{9}$$

The heat kernel is a solution of the *heat equation* on the line, $\dfrac{\partial H}{\partial t} = \dfrac{\partial^2 H}{\partial x^2}.$

Exercise 5 *Show that the Gaussian $\{K_\delta\}$, Féjer $\{F_R\}$, Poisson $\{P_y\}$, and heat kernels $\{H_t\}$, generate approximations of the identity on $\mathbb{R}$, as $\delta \to 0$, $R \to \infty$, $y \to 0$, and $t \to 0$. Assume known that $\int_{\mathbb{R}} \frac{1-\cos x}{x^2} dx = \pi$, and $\int_{\mathbb{R}} \frac{1}{1+x^2} dx = \pi$.*

Dirichlet kernel on $\mathbb{R}$:

$$D_R(x) = \int_{-R}^{R} e^{2\pi i x \xi} d\xi = \frac{\sin(2\pi R x)}{\pi x}. \tag{10}$$

Denote by $S_R f$ the "partial Fourier integral", $S_R(f) = \int_{-R}^{R} \hat{f}(\xi) 2^{2\pi i x \xi} d\xi$. Then $S_R f = D_R * f$.

Exercise 6 *The Dirichlet kernel is not integrable, hence the family $\{D_R\}$ does not generate an approximation of the identity (this is the origin of many of the convergence problems for Fourier series and integrals). Show that*

$$F_R(x) = \frac{1}{R}\int_0^R D_t(x)dt.$$

Conjugate Poisson kernel on $\mathbb{R}$:

$$Q_y(x) = \frac{x}{\pi(x^2+y^2)}, \quad y > 0. \tag{11}$$

Exercise 7 *Show that the conjugate Poisson kernel is not integrable either. The conjugate Poisson kernel satisfies Laplace's equation also, $\Delta Q = 0$. Also notice that $P_y(x) + iQ_y(x) = 1/(\pi i z)$, where $z = x + iy$.*

2.3 Time/Frequency Dictionary

The Fourier transform is a *linear transformation*. It interacts very nicely with *translations*, *modulations*, and *dilations*. We already mentioned that the Fourier transform converts differentiation into polynomial multiplication, and convolution into product. These properties are extremely useful.

Here is a *time/frequency dictionary* certainly valid in $\mathcal{S}(\mathbb{R})$,

	TIME	FREQUENCY
(a)	linear properties $af + bg$	linear properties $a\hat{f} + b\hat{g}$
(b)	translation $\tau_h f(x) = f(x-h)$	modulation $(\tau_h f)^\wedge(\xi) = e^{-2\pi i h\xi}\hat{f}(\xi) = M_{-h}\hat{f}(\xi)$
(c)	modulation $M_h f(x) = e^{2\pi i h x}f(x)$	translation $(e^{2\pi i h x}f)^\wedge(\xi) = \hat{f}(\xi - h) = \tau_h\hat{f}(\xi)$
(d)	dilation $f_s(x) = s^{-1}f(s^{-1}x)$	dilation $(f_s)^\wedge(\xi) = \hat{f}(s\xi)$
(e)	reflection $\tilde{f}(x) = f(-x)$	reflection $(\tilde{f})^\wedge(\xi) = -\hat{f}(-\xi)$
(f)	conjugate $\overline{f}(x) = \overline{f(x)}$	conjugate reflection $(\overline{f})^\wedge(\xi) = \overline{\hat{f}(-\xi)} = \overline{\check{f}}(\xi)$
(g)	derivative $f'(x)$	polynomial $(f')^\wedge(\xi) = 2\pi i\xi\hat{f}(\xi)$
(h)	polynomial $-2\pi i x f(x)$	derivative $[-2\pi i x f(x)]^\wedge(\xi) = \frac{d}{d\xi}\hat{f}(\xi)$
(i)	convolution $f * g(x) = \int f(x-y)g(y)dy$	product $(f*g)^\wedge(\xi) = \hat{f}(\xi)\hat{g}(\xi)$

Properties (b)-(e) are *symmetry* properties, or *group invariance* properties, which are fundamental in the theory. One can develop a Fourier theory on groups. The role of the trigonometric functions will now be played by the *characters* in the group. We will not pursue this here.

Exercise 8 *Check identity* (5) *using the dictionary above.*

Example 3 *We can use these properties to prove Plancherel's identity. Set* $\hat{g} = \overline{f}$ *in the multiplication formula* (4). *Then, by the analogue of property* (f) *for the inverse Fourier transform,* $g = (\overline{f})^{\vee} = \overline{\hat{f}}$, *therefore,*

$$\int |f|^2 = \int f\overline{f} = \int f\hat{g} = \int \hat{f}g = \int \hat{f}\overline{\hat{f}} = \int |\hat{f}|^2.$$

Let us prove Property (h) which does not have an immediate analogue in the Fourier series theory. The rest are left as exercises.

Proof of (h): This can be immediately believed by differentiating under the integral sign. We will justify the interchange of the limit and the integral.

We will check that $\hat{f}$ is differentiable, and that its derivative coincides with the Fourier transform of $-2\pi ixf(x)$, both in one stroke! Consider

$$\frac{\hat{f}(\xi+h)-\hat{f}(\xi)}{h} - [-2\pi ixf(x)]^{\wedge}(\xi) = \int_{\mathbb{R}} f(x)e^{-2\pi ix\xi}\left(\frac{e^{-2\pi ixh}-1}{h} + 2\pi ix\right)dx. \tag{12}$$

Suffices to check that we can make the absolute value of this quantity smaller than a constant multiple of $\epsilon > 0$, for any $\epsilon > 0$. We can bound separately the contributions for large x, and for small x. For all $N > 0$,

$$|(12)| \leq \underbrace{\int_{|x|\leq N} |f(x)|\left|\frac{e^{-2\pi ixh}-1}{h} + 2\pi ix\right| dx}_{(A)} + \underbrace{2\pi\int_{|x|>N} |x||f(x)|\left|\frac{e^{-2\pi ixh}-1}{2\pi xh} + 1\right| dx}_{(B)}.$$

Since $f \in \mathcal{S}(\mathbb{R})$ so is $xf(x)$, f is bounded, say by $M > 0$, and given $\epsilon > 0$, there exists $N > 0$ so that

$$\int_{|x|\geq N} |xf(x)|dx \leq \epsilon.$$

Moreover, for $|x| \leq N$, there exists h_0 so that for all $h < h_0$

$$\left|\frac{e^{-2\pi ixh}-1}{h} + 2\pi ix\right| \leq \frac{\epsilon}{N}.$$

Also remember that for all real θ, $|e^{i\theta}-1| \leq |\theta|$, in particular, $\left|\frac{e^{i\theta}-1}{\theta} + 1\right| \leq 2$. All these facts together imply that $(A) \leq 2M\epsilon$, and $(B) \leq 4\pi\epsilon$. Hence there exists a $C > 0$ such that for all $\epsilon > 0$ there exists $h_0 > 0$ such that for all $h < h_0$,

$$\left| \frac{\hat{f}(\xi + h) - \hat{f}(\xi)}{h} - \int_{\mathbb{R}} -2\pi i x f(x) e^{-2\pi i x \xi} dx \right| < C\epsilon.$$

We conclude that, $\dfrac{d}{d\xi}\hat{f}(\xi) = [-2\pi i x f(x)]^{\wedge}(\xi)$. $\diamond$

In view of these properties, *smoothness of the function is correlated with fast decay of the Fourier transform at infinity.* In particular, band-limited functions (compactly supported Fourier transform) have the most dramatic decay at infinity, hence they correspond to infinitely differentiable functions (C^∞). There is also a whole lot that can be said about decay properties of a function at infinity, and the analyticity properties of its Fourier transform, such results go under the generic name of Paley-Wiener Theorems (see [RS, ch. IX.3]).

Example 4 *Find a function f on $\mathbb{R}$ that satisfies the differential equation*

$$f^{(3)}(x) + 3f''(x) - 2f'(x) - 6f(x) = e^{-\pi x^2}.$$

Taking the Fourier transform on both sides of the identity and using the time/frequency dictionary, we obtain

$$\hat{f}(\xi)[(2\pi i\xi)^3 + 3(2\pi i\xi)^2 - 2(2\pi i\xi) - 6] = e^{-\pi\xi^2}. \tag{13}$$

Let $p(t) = t^3 + 3t^2 - 2t - 6 = (t+3)(t^2-2)$. Then the polynomial inside the brackets in (13), *is $q(\xi) = p(2\pi i\xi)$. Note that $q(t)$ has no real zeros, hence we can solve safely for $\hat{f}$, and then use the Fourier inversion formula in $\mathcal{S}$,*

$$\hat{f}(\xi) = \frac{e^{-\pi\xi^2}}{q(\xi)} \Rightarrow f(x) = \left(\frac{e^{-\pi\xi^2}}{q(\xi)}\right)^{\vee}(x).$$

In general, let $P(x) = \sum_{k=0}^{n} a_k x^k$ be a polynomial of degree n with constant complex coefficients. If $P(x)$ has no real zeros, and $u \in \mathcal{S}(\mathbb{R})$, then the linear differential equation

$$P(D)f = \sum_{k=0}^{n} a_k D^k f = u$$

has a solution given by

$$f = \big(\hat{u}(\xi)P(\xi)^{-1}\big)^{\vee}.$$

Note that since $P(\xi)$ has no real zeros the function $\hat{u}(\xi)P(\xi)^{-1}$ is in the Schwartz class.

2.4 Beyond $\mathcal{S}$

These properties hold in a larger class of functions, it will suffice for most of our illustrations to consider *functions of moderate decrease*, for which we could still be rigorous. Such functions have the property that they are continuous and there exists a constant $A > 0$ so that

$$|f(x)| \leq \frac{A}{1+x^2} \quad \text{for all} \quad x \in \mathbb{R}.$$

Example 5 *All Schwartz functions are functions of moderate decrease. However there are functions, like*

$$\frac{1}{1+|x|^n}, \quad n \geq 2; \quad \text{and} \quad e^{-a|x|};$$

which are of moderate decrease, although they are not in $\mathcal{S}(\mathbb{R})$.

In this space one has a satisfactory theory of integration. For f a function of moderate decrease we can define

$$\int_{-\infty}^{\infty} f(x)dx = \lim_{N\to\infty} \int_{-N}^{N} f(x)dx.$$

The Fourier transform is well defined for functions of moderate decrease. The inversion formula and Plancherel hold when f and $\hat{f}$ are both of moderate decrease. See [StSh, Ch. 5, Sec 1.1 & 1.7] for more details.

Exercise 9 *Neither the Poisson* (8) *or Fejér* (7) *kernels, belong to* $\mathcal{S}(\mathbb{R})$, *however they are functions of moderate decrease. Show that the Fourier transform of the Poisson kernel is* $e^{-2\pi|\xi|y}$. *Show that the Fourier transform of the Fejér kernel is* $\left(1 - \frac{|\xi|}{R}\right)$ *if* $|\xi| \leq R$, *zero otherwise.*

Exercise 10 *Show that the convolution of two functions of moderate decrease is a function of moderate decrease. See* [StSh, Ch 5, ex 7].

2.4.1 Tempered distributions and $L^p(\mathbb{R})$

One can extend most of the basic operations (translation, dilation, differentiation) to a larger class of objects, the space of *tempered distributions* $\mathcal{S}'(\mathbb{R})$. These are the *continuous linear transformations*[2] $T : \mathcal{S} \to \mathbb{C}$. Given $T \in \mathcal{S}'(\mathbb{R})$, we define its translate by h, $\tau_h T$, its dilation by s, T_s, and its derivative T', by the following formulae,

$$T_s(\phi) = T\big((1/s)\phi_{1/s}\big), \quad \tau_h T(\phi) = T(\tau_{-h}\phi), \quad T'(\phi) = -T(\phi'), \quad \text{for all } \phi \in \mathcal{S}(\mathbb{R}). \tag{14}$$

[2]Here we are brushing under the rug a lot! The word "continuous" means there is an underlying *topology* in $\mathcal{S}$. In that case, linear transformations are continuous, if given a sequence $\{\phi_n\}$ convergent in the topology of $\mathcal{S}$ to $\phi \in \mathcal{S}$, then $T(\phi_n)$ converges to $T(\phi)$. Meaningful extensions of the basic operations, and the Fourier transform, will have to be nicely behaved with respect to the corresponding topology. See [StWe, ch. 1] for a careful introduction to distribution theory. We will not define the topology, nor check continuity issues in these notes.

The Fourier transform can also be extended to be a bijection on $\mathcal{S}'(\mathbb{R})$. The Fourier transform of such T, denoted by $\hat{T}$, is the tempered distribution defined by

$$\hat{T}(\phi) = T(\hat{\phi}), \quad \text{for all } \phi \in \mathcal{S}(\mathbb{R}).$$

Similarly can define the inverse Fourier transform, $\check{T}(\phi) = T(\check{\phi}), \quad$ for all $\phi \in \mathcal{S}(\mathbb{R})$.

Example 6 *All* locally integrable[3] *functions f define a tempered distribution by integration,*

$$T_f(\phi) = \int_{\mathbb{R}} f(x)\phi(x)dx. \tag{15}$$

In particular functions in the Schwartz class are locally integrable, and so are functions in the classical *Lebesgue spaces*, $L^p(\mathbb{R})$, $1 \leq p < \infty$, which consist of those functions such that $\int_{\mathbb{R}} |f|^p < \infty$. Here we are entering the territory of *Lebesgue integration*. It suffices to say that Lebesgue integration is to Riemann integration what real numbers are to rational numbers. With such integration theory, all these spaces are *normed*, with norm given by,

$$\|f\|_p = (\int_{\mathbb{R}} |f(x)|^p dx)^{1/p},$$

and *complete*[4]. That is, they are *Banach spaces*. The space of square integrable functions on the line, $L^2(\mathbb{R})$, is a *Hilbert space*, with *inner product* given by

$$\langle f, g\rangle = \int_{\mathbb{R}} f(x)\overline{g(x)}dx.$$

The Schwartz class is *dense* in all $L^p(\mathbb{R})$. That is, we can approximate any function $f \in L^p(\mathbb{R})$ by functions $\phi_n \in \mathcal{S}(\mathbb{R})$ such that

$$\lim_{n\to\infty} \|f - \phi_n\|_p = 0.$$

We will say that an identity, $f = g$, or a limit, $\lim f_t = f$, holds in the L^p-sense, if,

$$\|f - g\|_p = 0, \quad \text{or} \quad \lim \|f_t - f\|_p = 0.$$

Notice that the operations in (14) are defined so that they coincide with the usual dilation, translation and differentiation operations when we consider a function to be a tempered distribution.

Exercise 11 *Check that for any function $f \in \mathcal{S}(\mathbb{R})$,*

$$(T_f)_s = T_{f_s}, \quad \tau_h T_f = T_{\tau_h f}, \quad (T_f)' = T_{f'}, \quad \textit{and} \quad \hat{T}_f = T_{\hat{f}}.$$

---

[3]This means that the for all compact intervals $[a, b]$, $\int_a^b |f(x)|dx$ is finite.

[4]That is, every Cauchy sequence in the space converges to an element in the space.

But there are more than functions in $\mathcal{S}'(\mathbb{R})$. Borel measures (for those of you who know them) can be viewed as tempered distributions integrating against them.

Example 7 *The so-called* delta function *(point-mass measure!) is a tempered distribution, but not a function in the usual sense! It is defined to be*

$$\delta(\phi) = \phi(0),$$

and we can compute its Fourier transform,

$$\hat{\delta}(\phi) = \hat{\phi}(0) = \int \phi(x)dx,$$

that is, $\hat{\delta}$ corresponds to the tempered distribution given by the function identically equal to one, see (15). *We write $\hat{\delta} = 1$, but we mean $\hat{\delta} = T_1$.*

This example illustrates a time/frequency principle:

> *The smaller the support of the function, the larger the support of its Fourier transform, and viceversa.*

This will be made precise by the *Heisenberg Uncertainty Principle* in Section 2.6.

Exercise 12 *Compute the derivatives of the delta function. These tempered distributions are not in any of the categories described before! Check that their Fourier transforms $(D^k(\delta))^\wedge$ can be identified with the polynomials $(2\pi ix)^k$. Check that δ coincides, in the sense of distributions, with the derivative of the* heavy-side function $H(x) = 0$ *if* $x \leq 0$, $H(x) = 1$ *if* $x > 0$.

Here is a table of how the Fourier transform acts on these spaces

ϕ	$\to$	$\hat{\phi}$
$\mathcal{S}$	unitary bijection	$\mathcal{S}$
L^1	bounded map $\lVert\hat{f}\rVert_\infty \leq \lVert f\rVert_1$ (Riemann-Lebesgue Lemma)	$C_0 \subset L^\infty$
L^2	isometry $\lVert\hat{f}\rVert_2 = \lVert f\rVert_2$ (Plancherel)	L^2
L^p $1 < p < 2$	bounded map $\lVert\hat{f}\rVert_q \leq C_p\lVert f\rVert_p$ (Hausdorff-Young's inequality)	L^q, $\frac{1}{p} + \frac{1}{q} = 1$
$\mathcal{S}'$	bijection	$\mathcal{S}'$

$C_0(\mathbb{R})$ denotes the continuous functions vanishing at infinity, which is a subset of $L^\infty(\mathbb{R})$, the space of *essentially bounded*[5] functions on the line. The Riemann-Lebesgue lemma is a strong statement, the Fourier transform of an integrable function ($f \in L^1(\mathbb{R})$), is a continuous function such that $\lim_{\xi\to\pm\infty} \hat{f}(\xi) = 0$. However the Fourier transform is not a surjective map from $L^1(\mathbb{R})$ to $C_0(\mathbb{R})$ (see [Kr, Prop. 2.3.15 p. 112)].

Exercise 13 *Check the Riemann-Lebesgue Lemma for continuous functions f of moderate decrease. See Exercise* 5 *in* [StSh, Ch. 5].

The Fourier transform is well defined in $L^1(\mathbb{R})$, but since $\hat{f}$ is not necessarily integrable, one has difficulties with the inversion formula. This is why we have chosen to present the theory on $\mathcal{S}(\mathbb{R})$. Since $\mathcal{S}(\mathbb{R})$ is dense in both $L^1(\mathbb{R})$ and $L^2(\mathbb{R})$, one can extend by continuity the Fourier transform to these spaces. In $L^2(\mathbb{R})$, Plancherel will be transferred by continuity as well. Are these extensions the same on $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$? The answer is yes. Notice that in $\mathbb{R}$, we do not have the nice ladder structure for L^p-spaces that we had on the circle, in particular $L^2(\mathbb{R})$ is not a subset of $L^1(\mathbb{R})$, nor is $L^1(\mathbb{R})$ a subset of $L^2(\mathbb{R})$.

The Fourier transform does not map $L^p(\mathbb{R})$ for $p > 2$ into a nice space. What is so special about $L^p(\mathbb{R})$, $1 < p < 2$? It turns out that in that range it is true that $L^p(\mathbb{R}) \subset L^1(\mathbb{R}) + L^2(\mathbb{R})$, and one can then use the theory in $L^1(\mathbb{R})$ and $L^2(\mathbb{R})$. We will come back to this issue in the fourth lecture when we will discuss interpolation.

Exercise 14 *Given $1 \leq p < q \leq \infty$. Find a function in $L^q(\mathbb{R})$ that is not in $L^p(\mathbb{R})$. Find a function that is in $L^p(\mathbb{R})$ but is not in $L^q(\mathbb{R})$.*

Exercise 15 *The conjugate Poisson kernel $Q_y(x)$, see* (11), *is in $L^2(\mathbb{R})$ for each $y > 0$. Check that $\hat{Q}_y(\xi) = -i\,\mathrm{sgn}(\xi)e^{-2\pi|y\xi|}$.*

2.5 Poisson Summation Formula

Given a function on the line, how can we construct a periodic function of period 1?

There are two approaches, one is called *periodization*, the other uses a Fourier series whose Fourier coefficients are given by the Fourier transform evaluated at the integers. The Poisson Summation Formula gives conditions under which these two methods coincide.

The periodization of a function f on the line is constructed by summing over its integer translates,

$$f^p(x) = \sum_{n\in\mathbb{Z}} f(x+n) = \lim_{N\to\infty} \sum_{|n|\leq N} f(x+n),$$

where the series is supposed to converge in the sense of symmetric partial sums.

Example 8 *If $f \in L^1(0,1)$ and is defined to be zero elsewhere, then f^p is just its periodic extension to the whole line.*

[5] This are functions that are bounded except on a set of *measure zero*. It is a Banach space with norm given by, $\|f\|_\infty = \mathrm{esssup}_{x\in\mathbb{R}}|f(x)|$. Think of $L^\infty(\mathbb{R})$ as the space of bounded functions on the line.

Example 9 *If H_t is the* heat kernel *on the line, defined by* (9)*, then its periodization is the periodic heat kernel* $H_t^p(x) = (4\pi t)^{-1/2}\sum_{n\in\mathbb{Z}} e^{-(x+n)^2/4t}$.

Exercise 16 *Show that the periodization of the Poisson, Dirichlet, Fejér, and conjugate Poisson kernels on the line, coincide with the periodic Poisson, Dirichlet, Fejér, and conjugate Poisson kernels, most of them introduced in the first part of the course. In particular see* [StSh, Ch 5, Sec. 3.2-3.3 & Exs. 9, 14, 16].

Theorem 3 (Poisson Summation Formula) *If $f \in \mathcal{S}(\mathbb{R})$ then*

$$\sum_{n\in\mathbb{Z}} f(x+n) = f^p(x) = \sum_{n\in\mathbb{Z}} \hat{f}(n)e^{2\pi i n x}, \tag{16}$$

in particular, setting $x=0$,

$$\sum_{n\in\mathbb{Z}} f(n) = \sum_{n\in\mathbb{Z}} \hat{f}(n).$$

Proof: It suffices to check that the right and left sides of (16) have the same Fourier coefficients. The n^{th} Fourier coefficient on the right-hand-side is $\hat{f}(n)$. Computing for the left-hand-side we get,

$$\begin{aligned}
\int_0^1 \left(\sum_{m\in\mathbb{Z}} f(x+m)\right) e^{-2\pi i n x} dx &= \sum_{m\in\mathbb{Z}} \int_0^1 f(x+m)e^{-2\pi i n x} dx \\
&= \sum_{m\in\mathbb{Z}} \int_m^{m+1} f(y)e^{-2\pi i n y} dy \\
&= \int_{-\infty}^{\infty} f(y)e^{-2\pi i n y} dy \\
&= \hat{f}(n).
\end{aligned}$$

We are entitled to interchange the sum and the integral in the first step since f is rapidly decreasing. $\diamond$

Example 10 *If H_t is the heat kernel, then all conditions in the Poisson Summation Formula are satisfied, furthermore $\hat{H}_t(n) = e^{-4\pi^2 t n^2}$, hence the periodic heat kernel has a natural Fourier series in terms of separated solutions of the heat equation,*

$$H_t^p(x) = (4\pi t)^{-1/2}\sum_{n\in\mathbb{Z}} e^{-(x+n)^2/4t} = \sum_{n\in\mathbb{Z}} e^{-4\pi^2 t n^2} e^{2\pi i n x}.$$

One of the most useful applications of the Poisson Summation Formula is the *Shannon Sampling Formula.* It states that band-limited functions (functions whose Fourier transform has compact support) can be recovered by appropriate samplings as coefficients of series of translated sinc functions. Remember that

$$\operatorname{sinc}(x) = \frac{\sin(\pi x)}{\pi x}.$$

You will explore this formula in the problem session.

Exercise 17 *Do exercises* 2 *and* 20*, in Ch* 5 *in* [StSh].

2.6 Heisenberg's Uncertainty Principle

It is impossible to find a function which is simultaneously well localized in time and frequency. Actually this can be carried one step further, one cannot construct a function which is compactly supported and band-limited.

Theorem 4 (Heisenberg's Uncertainty Principle) *Suppose $\psi \in \mathcal{S}(\mathbb{R})$, and it is normalized in $L^2(\mathbb{R})$, i.e. $\|\psi\|_2 = 1$. Then,*

$$\int_{\mathbb{R}} x^2 |\psi(x)|^2 dx \int_{\mathbb{R}} \xi^2 |\hat{\psi}(\xi)|^2 d\xi \geq \frac{1}{16\pi^2}.$$

Proof: By hypothesis,

$$1 = \int_{\mathbb{R}} |\psi(x)|^2 dx.$$

We want to integrate by parts, setting $u = |\psi(x)|^2$, and $dv = dx$, and take advantage of the fast decay of ψ to get rid of the boundary terms. Note that

$$\frac{d}{dx}|\psi|^2 = \frac{d}{dx}\psi\overline{\psi} = \overline{\psi}\psi' + \overline{\overline{\psi}\psi'} = 2\mathrm{Re}(\overline{\psi}\psi').$$

Also remember that $\mathrm{Re} z \leq |z|$, and that $|\int f| \leq \int |f|$. Hence

$$1 = 2\int_{\mathbb{R}} x\mathrm{Re}\big(\overline{\psi(x)}\psi'(x)\big)\, dx \leq \int_{\mathbb{R}} |x||\psi(x)||\psi'(x)|.$$

Apply now the celebrated Cauchy-Schwarz inequality[6] to the right-hand side, and get,

$$1 \leq 2\left(\int_{\mathbb{R}} |x\psi(x)|^2 dx\right)^{1/2} \left(\int_{\mathbb{R}} |\psi'(x)|^2 dx\right)^{1/2}.$$

This gives us the theorem, after observing that by Plancherel and the time/frequency dictionary, in particular (g),

$$\int_{\mathbb{R}} |\psi'(x)|^2 dx = 4\pi^2 \int_{\mathbb{R}} |\xi\hat{\psi}(\xi)|^2 dx.$$

$\diamond$

Exercise 18 *Check that equality holds if and only if $\psi(x) = \sqrt{2B/\pi}e^{-Bx^2}$, for $B > 0$.*

[6] The Cauchy-Schwarz inequality valid on inner-product vector spaces, says that

$$|\langle u, v\rangle| \leq \langle u, u\rangle^{1/2}\langle v, v\rangle^{1/2}.$$

Exercise 19 *Check that under the conditions of the above theorem,*

$$\int_{\mathbb{R}} (x - x_0)^2 |\psi(x)|^2 dx \int_{\mathbb{R}} (\xi - \xi_0)^2 |\hat{\psi}(\xi)|^2 d\xi \geq \frac{1}{16\pi^2},$$

for all $x_0, \xi_0 \in \mathbb{R}$.

Exercise 20 *Justify the statement that one cannot find a function simultaneously supported in space and frequency. Suppose* f *is continuous on* $\mathbb{R}$. *Show that* f *and* $\hat{f}$ *cannot be both compactly supported, unless* $f = 0$. **Hint:** *Assume* f *is supported on* $[0, 1/2]$, *expand* f *in a Fourier series on* $[0, 1]$, *observe that* f *must be a trigonometric polynomial.*

The trigonometric functions are not in $L^2(\mathbb{R})$. However we can view them like distributions, and we can compute their Fourier transforms.

Exercise 21 *Check that the Fourier transform of* $e_\xi(x) = e^{2\pi i x\xi}$ *in the sense of distributions is the shifted* δ*-function,* δ_ξ, *defined to be* $\delta_\xi(\phi) = \phi(\xi)$.

This also illustrates the time/frequency localization principle.

Exercise 22 *Prove the analogues of the previous exercise in the finite dimensional context, and in the Fourier series context. Which vectors/sequences play the role of the* δ*-function?*

In general, if the support of a function is essentially localized on an interval of length d, then its Fourier transform will be essentially localized on an interval of length d^{-1}. This says that the support on the *phase* or *time/frequency plane* will be essentially a *rectangle* of area 1 and dimensions $d \times d^{-1}$.

Exercise 23 *Draw the phase planes for the trigonometric basis, and for the standard basis in finite dimensional space of dimension 8.*

3 From Fourier to Haar

In this lecture we will give a a brief survey on the windowed Fourier (Gabor) transform, and introduce the latest member of the family, wavelet analysis. We will then discuss more carefully the Haar basis.

3.1 Windowed Fourier transform

The continuous Fourier transform provides a tool for analyzing a function, but the exponentials can not be viewed as a "countable basis" anymore. The windowed Fourier transform addresses this problem.

How can we get an orthonormal basis in $L^2(\mathbb{R})$? A simple solution will be to split the line into unit segments, $[k, k+1]$, indexed by k, and on each segment use the periodic Fourier basis. Let χ_A denote the characteristic function of the set $A \in \mathbb{R}$,

$$\chi_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{otherwise} \end{cases}.$$

The functions

$$g_{n,k}(t) = e^{2\pi int}\chi_{[k,k+1]}(t), \quad n, k \in \mathbb{Z},$$

form an orthonormal basis for $L^2(\mathbb{R})$. They give us the so-called *windowed Fourier transform*,

$$Gf(n,k) = \langle f, g_{n,k}\rangle = \int_k^{k+1} f(t)e^{-2\pi int}dt,$$

and the corresponding reconstruction formula,

$$f(x) = \sum_{j,k} Gf(n,k)g_{n,k}(x).$$

However the sharp windows produce artifacts at the edges when dealing with numerical calculations. Smoother windows are desirable. The sharp windows $\chi_{[0,1]}$ and its translates, are replaced by a *smooth window* g and its integer translates

$$g_{n,k}(t) = g(t-k)e^{2\pi int}.$$

Gabor considered in 1946 systems of this type and proposed to utilize them in communication theory [Gab]. Notice that their Fourier transforms can be calculated using the time/frequency dictionary:

$$(g_{n,k})^{\wedge}(\xi) = \hat{g}(\xi - n)e^{-2\pi ik\xi} = \hat{g}_{-k,n}(\xi).$$

If g generates a Gabor basis, i.e. $\{g_{n,k}\}$ forms an orthonormal basis in $L^2(\mathbb{R})$, then so will $\hat{g}$.

Exercise 24 *Since $g = \chi_{[0,1]}$ generates an orthonormal basis, then so will*

$$\hat{g}(\xi) = (\chi_{[0,1]})^{\wedge}(\xi) = e^{-i\pi\xi}\frac{\sin(\pi\xi)}{(\pi\xi)} = e^{-i\pi\xi}\mathrm{sinc}(\xi).$$

This provides an example of a continuous window that fails to be differentiable at the origin.

The question now is what other functions $g \in L^2(\mathbb{R})$ will generate a Gabor basis? The limitations of the Gabor analysis are explained by the following result, which can be checked for the two examples we have exhibited,

Balian-Low Theorem: *For $g \in L^2(\mathbb{R})$, if $(g_{n,k})_{n,k\in\mathbb{Z}}$ is an orthonormal basis, then either*

$$\int t^2|g(t)|^2dt = \infty \quad \textit{or} \quad \int \xi^2|\hat{g}(\xi)|^2d\xi = \int |g'(t)|^2dt = \infty.$$

Proof: Can be found in [Dau, p.108, Thm 4.1.1.].

This theorem states that a Gabor window or bell cannot be simultaneously compactly supported and smooth. The first example, $g(x) = \chi_{[0,1]}(x)$, is perfectly localized in time but not in frequency, and the second example, $g(x) = e^{-i\pi x}\mathrm{sinc}(x)$, is the opposite. In particular the slow decay of the sinc function reflects its lack of smoothness. This phenomena is an incarnation of the *Heisenberg uncertainty principle*. However, if the exponentials are replaced by appropriate sines and cosines one can obtain a Gabor-type basis with smooth bell functions. These are the so-called *local sine and cosine bases*, described by Coifman and Meyer [CM], but first discovered by Malvar [Ma] (from Brazil).

There is a *continuous Gabor transform* as well, but the parameters are now real. Let g be a real and symmetric window, normalized so that $\|g\|_2 = 1$, let

$$g_{\xi,u}(t) = g(t-u)e^{2\pi i\xi t}, \qquad u, \xi \in \mathbb{R}.$$

The Gabor transform is then

$$Gf(\xi, u) = \int_{-\infty}^{\infty} f(t)g(t-u)e^{-2\pi i\xi t}dt.$$

The multiplication by the translated window localizes the Fourier integral in a neighborhood of u. The following *inversion formula* holds for $f \in L^2(\mathbb{R})$,

$$f(x) = \int_{\mathbb{R}}\int_{\mathbb{R}} Gf(u,\xi)g(x-u)e^{2\pi i x\xi}d\xi du.$$

3.2 Wavelet transform

Gabor bases give partial answers to the localization issues. A problem is that the size of the windows are fixed. Variable widths is the new ingredient added by wavelet analysis.

The wavelet transform involves translations (like Gabor basis) and scalings (instead of modulations). This provides a zooming mechanism which is behind the multiresolution structure of these bases.

The goal is to find functions $\psi \in L^2(\mathbb{R})$ so that the family

$$\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k), \qquad j,k \in \mathbb{Z},$$

forms an orthonormal basis of $L^2(\mathbb{R})$.

Exercise 25 *Assuming such a function exists, use the time/frequency dictionary to compute* $(\psi_{j,k})^\wedge(\xi)$.

The *orthogonal wavelet transform* is given by

$$Wf(j,k) = \langle f, \psi_{j,k}\rangle = \int_{\mathbb{R}} f(t)\overline{\psi_{j,k}(t)}dt,$$

and the reconstruction formula is,

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

The oldest example of a wavelet basis is provided by the *Haar basis* introduced by Haar in 1910 [Ha].

Example 11 *The* Haar function *on the unit interval is given by,*

$$h(t) = \chi_{[0,1/2)}(t) - \chi_{[1/2,1)}(t).$$

The family $\{h_{j,k}(t) = 2^{j/2}h(2^j t - k)\}_{j,k\in\mathbb{Z}}$ *is an orthonormal basis for* $L^2(\mathbb{R})$. *We will study this example in depth in the next section.*

Exercise 26 *Show that* $\{h_{j,k}\}$ *is an orthonormal set, and the functions have mean value zero. That is, show that* $\int h_{j,k} = 0$, $\|h_{j,k}\|_2 = 1$. *Also show that,*

$$\langle h_{j,k}, h_{j',k'}\rangle = \begin{cases} 1 & \text{if } j = j',\ k = k', \\ 0 & \text{otherwise.} \end{cases}$$

Wavelet theory can sometimes be thought as the search for smoother wavelets. Haar is perfectly localized in time but not in frequency.

Exercise 27 *Find the Fourier transform of the Haar function* $h(t)$.

Example 12 *At the other end of the spectrum one can find the* Shannon basis. *Let*

$$\hat{\psi}(\xi) = e^{2\pi i\xi}\chi_{\{\xi:1/2<|\xi|\leq 1\}}(\xi).$$

The family $\{\psi_{j,k}\}_{j,k\in\mathbb{Z}}$ *is an orthonormal basis for* $L^2(\mathbb{R})$.

The Shannon wavelets are perfectly localized in frequency but not in time.

Exercise 28 *Show that the Shannon functions,* $\{\psi_{j,k}\}$ *form an orthonormal set, and have mean value zero.* **Hint:** *Work on Fourier side using the Polarization formula* (6).

The goal is to find functions which are "simultaneously" localized in time and frequency (within the limits imposed by the Heisenberg principle). Compactly supported wavelets with arbitrary smoothness where constructed by I. Daubechies in a seminal paper in wavelet theory [Dau]. We can associate to most wavelets a sequence of numbers, a *filter*. Compactly supported wavelets have filters which are zero except for finitely many entries, such filters are called in the engineering community *finite impulse response filters* (FIR). The connections to filter bank theory and the possibility of implementing FIR filters opened the doors to widespread use of wavelets in applications. We will explore this connection in Section 4.2.

One can develop theory of wavelets in $\mathbb{R}^N$ or $\mathbb{C}^N$ (linear algebra!), the same way you built a finite Fourier theory last week. This is done in full detail in the book by Frazier [Fra]. This is what ends up being implemented. Not only do you have a FFT, it turns out you also have a *fast wavelet transform* which has been instrumental in the success of wavelets in the "real world". Let me just mention two of the most popular applications: the FBI fingerprint data base and retrieval, and the JPEG 2000 Standard for image compression are both based on wavelets.

As in the Fourier and Gabor cases, there is a *continuous wavelet transform* which can be traced back to the famous *Calderón reproducing formula.* In this case we have continuous translation and scaling parameters, $s \in \mathbb{R}^+$ (that is $s > 0$), $u \in \mathbb{R}$, and a family of *time/frequency atoms* is obtained by rescaling by s and shifting by u a normalized wavelet $\psi \in L^2(\mathbb{R})$ with zero average ($\int \psi = 0$).

Let

$$\psi_{s,u}(t) = \frac{1}{\sqrt{s}}\psi\left(\frac{t-u}{s}\right), \quad (\psi_{s,u})^\wedge(\xi) = \sqrt{s}e^{-2\pi iu\xi}\hat{\psi}(s\xi).$$

The continuous wavelet transform is then defined by,

$$Wf(s,u) = \langle f, \psi_{s,u}\rangle = \int_{\mathbb{R}} f(t)\overline{\psi_{s,u}(t)}dt.$$

If ψ is real valued, and it is localized near 0 with spread 1, then $\psi_{s,u}$ is localized near u with spread s. The wavelet transform measures the variation of f near u at scale s (in the orthonormal case, $u = k2^{-j}$ and $s = 2^{-j}$). As the scale s goes to zero (i.e. j goes to infinity), the decay of the wavelet coefficients characterizes the regularity of f near u (in the discrete case $k2^{-j}$). On the other hand, if $\hat{\psi}$ is localized near 0 with spread 1, then $(\psi_{s,u})^\wedge$ is localized near 0 with spread $1/s$. That is, the Heisenberg boxes or *phase-plane portrait* of the wavelets are rectangles of area one and dimensions $s \times 1/s$.

Under very mild assumptions on the wavelet ψ, we obtain a reconstruction formula. For any $f \in L^2(\mathbb{R})$,

$$f(t) = \frac{1}{C_\psi} \int_0^\infty \int_{-\infty}^{+\infty} Wf(s,u)\psi_{s,u}(t) \frac{du ds}{s^2},$$

provided that ψ satisfies *Calderón's admissibility condition* [Cal]

$$C_\psi = \int_0^\infty \frac{|\hat{\psi}(\xi)|^2}{\xi} d\xi < \infty.$$

3.3 Haar Analysis

In this section we will discuss carefully the Haar basis.

3.3.1 The dyadic intervals

Intervals of the form

$$I_{j,k} = [k2^{-j}, (k+1)2^{-j}), \quad \text{for integers } j, k,$$

are called *dyadic intervals*. The collection of all dyadic intervals is denoted by $\mathcal{D}$, and $\mathcal{D}_j$ denotes all dyadic intervals I of length 2^{-j}, also called the j-th generation. It is clear that each $\mathcal{D}_j$ provides a partition of the real line, and that

$$\mathcal{D} = \bigcup_{j \in \mathbb{Z}} \mathcal{D}_j.$$

Given two distinct intervals $I, J \in \mathcal{D}$, then either I and J are disjoint or one is contained in the other. Each dyadic interval I is in a unique generation $\mathcal{D}_j$, and there are exactly two subintervals of I in the next generation $\mathcal{D}_{j+1}$. These are the *children* of I, which we will denote I_r, I_l, the *right* and *left* child respectively. Clearly, $I = I_l \cup I_r$. We denote by $|I|$ the length of the interval I.

Exercise 29 *Show that if $I, J \in \mathcal{D}$, then $I \bigcap J = \emptyset$, or $I \subseteq J$, or $J \subset I$.*

Given $J \in \mathcal{D}$, we denote by $\mathcal{D}(J)$ the collection of dyadic intervals which are contained in J.

3.3.2 Haar Basis

Associated to each dyadic interval I, there is a *Haar function* defined by:

$$h_I(x) = \frac{1}{|I|^{1/2}} \left[\chi_{I_r}(x) - \chi_{I_l}(x) \right],$$

Exercise 30 *Let* $I_{j,k} = [k2^{-j}, (k+1)2^{-j})$*, show that*

$$h_{I_{j,k}}(x) = 2^{j/2}h(2^jx - k) = h_{j,k}(x), \quad \textit{where } \ h = h_{[0,1]}.$$

In Exercise 26 you were asked to check that $\{h_I\}_{I\in\mathcal{D}}$ is an orthonormal set in $L^2(\mathbb{R})$.

Exercise 31 *What would the Haar functions be in N-dimensional space ($N = 2^n$)? Are they an orthonormal basis in $\mathbb{R}^N$? If not, what are we missing?*

In the finite dimensional case, we can count the elements in the orthonormal set, and if that number coincides with the dimension, we know we have a basis. In infinite dimensional space we do not have that luxury. To prove Theorem 5 below, we must make sure the set is *complete.* It suffices to check that if there is an L^2-function orthogonal to all Haar functions, then it has no other choice than to be the zero function in $L^2(\mathbb{R})$. Or we can check directly that for all $f \in L^2(\mathbb{R})$, the following identity holds in the L^2-sense[7],

$$f = \sum_{I\in\mathcal{D}} \langle f, h_I\rangle h_I. \tag{17}$$

We will show that either argument boils down to checking some limit properties of the expectation operators defined in the next section. We will show that,

Theorem 5 *The Haar functions $\{h_I : I \in \mathcal{D}\}$ form an orthonormal basis in $L^2(\mathbb{R})$.*

3.3.3 The expectation and difference operators

We will introduce here two important operators that will help us understand the zooming properties of the Haar basis.

The *expectation* operators P_n, are nothing more than averages over dyadic intervals at generation n,

$$P_nf(x) = \frac{1}{|I|}\int_I f(t)\,dt,$$

where I is the unique interval of length 2^{-n} containing x.

The *difference* operators Q_n are given by,

$$Q_nf(x) = P_{n+1}f(x) - P_nf(x).$$

A telescoping series argument shows that

$$P_Nf(x) - P_Mf(x) = \sum_{M\le n<N} Q_nf(x). \tag{18}$$

[7]Recall that (17) holds in the L^2-sense if $\|f - \sum_{I\in\mathcal{D}}\langle f, h_I\rangle h_I\|_2 = 0$.

Lemma 1 *For $f \in L^2(\mathbb{R})$,*

$$Q_n f(x) = \sum_{I \in \mathcal{D}_n} \langle f, h_I \rangle h_I(x).$$

Proof: Here we are using the averaging property of integral averages on dyadic intervals, namely,

$$\frac{1}{|I|}\int_I f = \frac{1}{|I|}\left(\int_{I_r} f + \int_{I_l} f\right) = \frac{\frac{1}{|I_r|}\int_{I_r} f + \frac{1}{|I_l|}\int_{I_l} f}{2}. \tag{19}$$

More precisely, suppose $x \in I \in \mathcal{D}_n$, then $P_n f(x) = \frac{1}{|I|}\int_I f$, and

$$P_{n+1} f(x) = \begin{cases} \frac{1}{|I_r|}\int_{I_r} f & \text{if } x \in I_r \\ \frac{1}{|I_l|}\int_{I_l} f & \text{if } x \in I_l \end{cases}.$$

Hence,

$$Q_n f(x) = \begin{cases} \frac{1}{2}\left(\frac{1}{|I_r|}\int_{I_r} f - \frac{1}{|I_l|}\int_{I_l} f\right) & \text{if } x \in I_r \\ -\frac{1}{2}\left(\frac{1}{|I_r|}\int_{I_r} f - \frac{1}{|I_l|}\int_{I_l} f\right) & \text{if } x \in I_l \end{cases}.$$

Finally note that since $|I| = 2|I_r| = 2|I_l|$, then

$$\frac{1}{2}\left(\frac{1}{|I_r|}\int_{I_r} f - \frac{1}{|I_l|}\int_{I_l} f\right) = \frac{1}{|I|^{1/2}}\langle f, h_I \rangle. \tag{20}$$

Since $h_I(x) = |I|^{-1/2}$ if $x \in I_r$, and $h_I(x) = -|I|^{-1/2}$ if $x \in I_l$, we conclude that,

$$Q_n f(x) = \langle f, h_I \rangle h_I(x) \quad \text{for } x \in I \in \mathcal{D}_n.$$

$\diamond$

3.3.4 Completeness of the Haar system

The Haar system of functions (wavelets) is complete if, for $f \in L^2(\mathbb{R})$, we have

$$f(x) = \sum_{I \in \mathcal{D}} \langle f, h_I \rangle h_I(x).$$

By Lemma 1 and (18), the right-hand side is equal to

$$\lim_{M,N \to \infty} \sum_{-M \le n < N} Q_n f(x) = \lim_{N \to \infty} P_N f(x) - \lim_{M \to -\infty} P_M f(x),$$

where all equalities are in the L^2-sense. We will be done if we can show the following:

Theorem 6 *For $f \in L^2(\mathbb{R})$,*

$$\lim_{M \to -\infty} \|P_M f\|_2 = 0, \tag{21}$$

$$\lim_{N \to \infty} \|P_N f - f\|_2 = 0. \tag{22}$$

Exercise 32 *Use Theorem 6 to show that, if $f \in L^2(\mathbb{R})$ is orthogonal to all Haar functions, then f must be zero in L^2. The function $f(x) = 1$ is orthogonal to all Haar functions, does this contradict the previous statement?*

Note that (22) says that given $x \in \mathbb{R}$, the averages of the function over the "tower" of dyadic intervals $\{I_n\}$ shrinking to $\{x\}$, $x \in I_n \in \mathcal{D}_n$, converges in L^2-sense to $f(x)$,

$$\lim_{n\to\infty} \frac{1}{|I_n|} \int_{I_n} f(t)\, dt = f(x).$$

It turns out that the convergence also holds almost everywhere (a.e). This is the content of the celebrated *Lebesgue Differentiation Theorem.*

Exercise 33 *Check the Lebesgue Differentiation Theorem for continuous functions, and that in this case the pointwise convergence holds everywhere. That is, for all $x \in \mathbb{R}$,*

$$\lim_{[a,b]\to\{x\}} \frac{1}{b-a} \int_a^b f(t)\, dt = f(x). \tag{23}$$

Note that we are not restricting to dyadic intervals in this case.

Theorem (6) will be a consequence of the following lemmata,

Lemma 2 *For any function $f \in L^2(\mathbb{R})$, and integer n, we have*

$$\|P_n f\|_2 \le \|f\|_2.$$

Lemma 3 *If g is continuous and has compact support on the interval $[-K, K]$, then Theorem 6 holds.*

Lemma 4 *The continuous functions with compact support are dense in $L^2(\mathbb{R})$. More precisely, given $f \in L^2(\mathbb{R})$, for any $\epsilon > 0$ there exist functions f and g, such that $f = g + h$, where g is a continuous function with compact support on an interval $[-K, K]$, and $h \in L^2(\mathbb{R})$ with small L^2-norm, $\|h\|_2 < \epsilon$.*

Proof of Theorem 6: By Lemma 4, given $\epsilon > 0$ we can decompose $f = g + h$, where g is continuous with compact support on $[-K, K]$, and $h \in L^2(\mathbb{R})$ with $\|h\|_2 < \epsilon/4$.

By Lemma 3, we can choose N large enough, so that for all $n > N$,

$$\|P_{-n} g\|_2 \le \epsilon/2.$$

Now apply the triangle inequality, and use Lemma 2 to conclude that

$$\|P_{-n} f\|_2 \le \|P_{-n} g\|_2 + \|P_{-n} h\|_2 \le \frac{\epsilon}{2} + \|h\|_2 \le \epsilon.$$

This shows (21).

By Lemma 3, we can choose N large enough, so that for all $n > N$,

$$\|P_n g - g\|_2 \leq \epsilon/2.$$

Now apply the triangle inequality, and use Lemma 2 to conclude that

$$\|P_n f - f\|_2 \leq \|P_n g - g\|_2 + \|P_n h - h\|_2 \leq \frac{\epsilon}{2} + 2\|h\|_2 \leq \epsilon.$$

This shows (22). $\diamond$

We have used twice a very important principle in functional analysis: *If a sequence of linear operators is uniformly bounded in a Banach space, and the the sequence converges to a bounded operator on a dense subset of the Banach space, then it converges in the whole space!!* This principle goes under the name *Uniform Boundedness Principle* or *Banach-Steinhaus Theorem.*

In our case the Banach space is $L^2(\mathbb{R})$, the dense subset is the set of continuous functions with compact support, the linear operators are P_n, and the uniform bounds are provided by Lemma 2. In one case the operators converge to the zero operator (as $n \to -\infty$), and in the other case they converge to the identity operator (as $n \to \infty$).

Exercise 34 *Show that the set $\{h_I : I \in \mathcal{D}([0,1])\}$ is not a complete set in $L^2([0,1])$. What are we missing? Can you complete the set?*

Proof of Lemma 2: For $x \in I \in \mathcal{D}_n$,

$$|P_n f(x)|^2 = \left|\frac{1}{|I|}\int_I f(t)dt\right|^2 \leq \frac{1}{|I|^2}\left(\int_I 1^2\,dt\right)\left(\int_I |f(t)|^2 dt\right) = \frac{1}{|I|}\int_I |f(t)|^2 dt,$$

the inequality is a consequence of the Cauchy-Schwarz inequality.

Now integrate over the interval I, to obtain

$$\int_I |P_n f(x)|^2 dx \leq \int_I |f(t)|^2 dt,$$

and add over all intervals in $\mathcal{D}_n$ (this is a disjoint family that covers the whole line!),

$$\int_{\mathbb{R}} |P_n f(x)|^2 dx = \sum_{I\in\mathcal{D}_n}\int_I |P_n f(x)|^2 dx \leq \sum_{I\in\mathcal{D}_n}\int_I |f(t)|^2 dt = \int_{\mathbb{R}} |f(t)|^2 dt.$$

The lemma is proven. $\diamond$

Proof of Lemma 3: The function g is continuous and has support on the interval $[-K, K]$. If n is large enough so that $K < 2^n$ and $x \in [0, 2^n] \in \mathcal{D}_{-n}$, then

$$\begin{aligned} |P_{-n} g(x)| &\leq \frac{1}{2^n}\int_0^K |g(t)|dt \\ &\leq \frac{1}{2^n}\left(\int_0^K 1^2 dt\right)^{1/2}\left(\int_0^K |g(t)|^2 dt\right)^{1/2} = \frac{1}{2^n}\sqrt{K}\|g\|_2, \end{aligned}$$

where the last inequality is another application of the Cauchy-Schwartz inequality. The same inequality holds for $x < 0$. If $|x| > 2^n$ then $P_{-n}g(x) = 0$.

We can now estimate the L^2-norm of $P_{-n}g$,

$$\|P_{-n}g\|_2^2 = \int_{-2^n}^{2^n} |P_{-n}g(x)|^2 dx \leq \frac{K}{2^{2n}} \|g\|_2^2 2^{n+1} = 2^{-n+1} K \|g\|_2^2.$$

By choosing N large enough, we can make $2^{-n+1}K\|g\|_2^2 < \epsilon^2$. That is, given $\epsilon > 0$, there is an $N > 0$ such that for all $n > N$,

$$\|P_{-n}g\|_2 \leq \epsilon.$$

This proves (21) for continuous functions with compact support.

We are assuming that g is continuous and it is supported on the compact interval $[-K, K]$. But this implies that g is uniformly continuous. That is, given $\epsilon > 0$ there exists $\delta > 0$ such that

$$|g(y) - g(x)| < \epsilon/\sqrt{4K} \quad \text{whenever} \quad |y - x| < \delta.$$

Now choose N large enough, so that $2^{-n} < \delta$ for all $n > N$. Given x, it is contained in a unique $I \in \mathcal{D}_n$, $|I| = 2^{-n} < \delta$, so $|y - x| \leq \delta$ for all $y \in I$, and

$$|P_n g(x) - g(x)| \leq \frac{1}{|I|} \int_I |g(y) - g(x)| dy \leq \frac{\epsilon}{\sqrt{4K}}.$$

Squaring and integrating we get

$$\int_{\mathbb{R}} |P_n g(x) - g(x)|^2 dx \leq \int_{-2K}^{2K} |P_n g(x) - g(x)|^2 dx + \int_{\{x \in \mathbb{R}: |x| > 2K\}} |P_n g(x)|^2 dx < \epsilon^2.$$

The first summand on the right-hand-side is smaller than ϵ^2 by previous considerations. The second summand is zero for n large enough, since for $|x| > 2K$, $P_n g(x)$ is the average over an interval $I \in \mathcal{D}_n$ which will be completely outside the support of g. For such x, $P_n g(x) = 0$ provided n is large enough. That is, given $\epsilon > 0$, there is an $N > 0$ such that for all $n > N$,

$$\|P_n g - g\|_2 \leq \epsilon.$$

This proves (22) for continuous functions with compact support.

$\diamond$

Proof of Lemma 4 This is an example of an approximation theorem in $L^2(\mathbb{R})$. How do we achieve this? We choose K large enough so that the "tail" of f has very small L^2- norm, that is $\|f\chi_{\{x \in \mathbb{R}: |x| > K\}}\|_2 \leq \epsilon/3$. Next we use the fact that on compact intervals, the continuous functions are dense in $L^2([-K, K])$ (you can think polynomials are dense, or trigonometric polynomials, by Weierstrass approximation theorem). Now choose g_1 continuous on $[-K, K]$ so that $\|f\chi_{[-K,K]} - g_1\|_2 \leq \epsilon/3$. It could happen that g_1 is continuous on $[-K, K]$, but when extended to be zero outside the interval, it is not continuous on the line. That can be fixed by giving yourself some margin at the endpoints, define g to coincide with g_1 on $[-K+\delta, K-\delta]$, to be zero outside $[-K,K]$, and connect with straight segments, so that g is continuous on $\mathbb{R}$. Finally, choose δ small enough so that $\|g_1 - g\|_2 \leq \epsilon/3$. Now let, $h = f - g = f\chi_{\{x\in\mathbb{R}:|x|>K\}} + f\chi_{[-K,K]} - g_1 + g_1 - g$, by triangle inequality we can check that

$$\|h\|_2 \leq \|f\chi_{\{x\in\mathbb{R}:|x|>K\}}\|_2 + \|f\chi_{[-K,K]} - g_1\|_2 + \|g_1 - g\|_2 \leq \epsilon.$$

◇

We have shown in Lemma 3 that the step functions can approximate continuous functions with compact support in the L^2-norm. In Lemma 4 we have argued that we can approximate L^2-functions by continuous functions with compact support. It means that (in the L^2-norm) we can approximate L^2-functions by step functions. Furthermore, we can choose the steps to be dyadic intervals of a fixed generation for any prescribed accuracy.

Exercise 35 *Show that you can approximate continuous functions with compact support with step functions in the uniform norm. Furthermore you can choose the steps to be dyadic intervals of a fixed generation for any prescribed accuracy. More precisely, show that given* $\epsilon > 0$ *there exists* $N > 0$ *such that for all* $n > N$*, and for all* $x \in \mathbb{R}$*,*

$$|P_n f(x) - f(x)| < \epsilon.$$

3.4 Haar vs Fourier

Two examples have been chosen to illustrate how the Haar basis can outperform the Fourier basis when it comes to deal with localized data, see [Kr, Sec. 7.4, p. 285].

The first example is a caricature of the problem: What could be the most localized "function" we could consider? The delta function. If we could find its Fourier series we will see that it has a very slowly decreasing tail that extends well beyond the highly localized support of the delta function. However when computing its Haar transform, is very localized, although it still has a tail, the tail decays faster.

Consider the following approximation of the δ-function

$$f_N(x) = 2^N \chi_{[0,2^{-N}]}.$$

Each of these functions have mass 1, and they converge in an appropriate sense to the δ-function.

Exercise 36 (a) *Compute the Fourier transform of* f_N*. Check that if we view* f_N *like a periodic function on* $[0,1]$*, and we compute its Fourier series, if its* M*-th partial Fourier sum is*

$$S_M(f_N)(x) = \sum_{|m|\leq M} \hat{f}(m) e^{2\pi i m x},$$

then

$$\|f_N - S_M(f_N)\|_2 \sim M^{-1/2}.$$

(b) *Compute the Haar coefficients of f_n. Check that the partial Haar sum,*

$$P_M(f_N)(x) = \sum_{j<M} Q_j(f_N)(x) = \sum_{|j|<M} Q_j(f_N)(x) + P_{-M+1},$$

has an exponential decay rate, that is,

$$\|f_N - P_M(f_N)\|_2 = 2^{-M/2}.$$

The localization properties of wavelets permit us to worry only about the wavelets which are near were the action occurs in the function. For example, if the function is constant on an interval, then the Haar coefficients of those wavelets supported in the interval vanish, because wavelets share the vanishing mean property that Haar functions have, that is $\int \psi = 0$. This is not the case with the trigonometric functions, whose support spreads all over.

Exercise 37 *Consider now the function f, defined on the real line,*

$$f(x) = \cos(\pi x)\chi_{[0,1]}(x).$$

Compute the Fourier transform of f and then reconstruct with the aid of the computer if necessary. Compute with the aid of the computer its Haar decomposition, and compare with the Fourier approximation.

In this example/exercise, had we consider the function $\cos(\pi x)\chi_{[-1,1]}(x)$ which goes over a full period of the cosine function, then the Fourier series will be the function itself, and no doubt will be the right way of doing the analysis. However, by chopping the function we force other Fourier coefficients to come into play.

3.4.1 Unconditional bases

The Haar basis happens to be a so-called *unconditional basis* for $L^p(\mathbb{R})$, $1 < p < \infty$. This basically means that we can approximate a function in the L^p-norm with finite linear combinations of Haar functions (basis). Furthermore the coefficients corresponding to h_I must be $\langle f, h_I\rangle$, and we can recover the L^p-norm of the function from knowledge about the *absolute value* of those coefficients, that is some formula involving only $|\langle f, h_I\rangle|$. No information about the sign/argument is necessary. In particular this means that if $f \in L^p$, and

$$f = \sum_{I\in\mathcal{D}} \langle f, h_I\rangle h_I,$$

then the new functions defined by

$$T_\sigma f = \sum_{I\in\mathcal{D}} \sigma_I \langle f, h_I\rangle h_I, \tag{24}$$

where $\sigma_I = \pm 1$ (or more generally $|\sigma_I| = 1$), should also be in L^p with norm comparable to that of f. There exist constants $c, C > 0$, such that for all choices of signs σ,

$$c\|f\|_p \leq \|T_\sigma f\|_p \leq C\|f\|_p. \tag{25}$$

For a given sequence $\sigma = \{\sigma_I\}$, the operator T_σ in (24) is called the *martingale transform*, and it is an example of a *constant Haar multipliers*. We will encounter these operators again in the last lecture.

Let us illustrate the above paragraphs in the case $p = 2$. We already know the Haar functions provide an orthonormal basis in $L^2(\mathbb{R})$, in particular, the Plancherel formula is valid:

$$\|f\|_2^2 = \sum_{I \in \mathcal{D}} |\langle f, h_I \rangle|^2.$$

To compute the L^2-norm of the function, we must add the squares of the absolute values of the Haar coefficients! It is clear that in this case,

$$\|T_\sigma f\|_2 = \|f\|_2,$$

hence (25) holds with $c = C = 1$.

It turns out that in the case of the Haar functions, one has the following norm equivalence,

$$c_p\|f\|_p \leq \|S^d(f)\|_p \leq C_p\|f\|_p. \tag{26}$$

Where the *dyadic square function* is defined to be

$$S^d(f)(x) = \left(\sum_{I \in \mathcal{D}} \frac{|\langle f, h_I \rangle|^2}{|I|} \chi_I(x) \right)^{1/2}. \tag{27}$$

This implies (25), because $S^d(f) = S^d(T_\sigma f)$. We can estimate the L^p-norm of f from the knowledge of the absolute values of its Haar coefficients alone.

Exercise 38 *Check that*

$$S^d(f)(x) = \left(\sum_j |Q_j f(x)|^2 \right)^{1/2}.$$

Check also that $\|S(f)\|_2 = \|f\|_2$.

In the case of general wavelets we will also have some *averaging* and *difference operators*, P_j and Q_j, and a corresponding *square function*. The same norm equivalence (26) will hold in $L^p(\mathbb{R})$. As it turns out, wavelet basis will provide unconditional basis for a whole zoo of functional spaces (Sobolev, Hölder, etc).

The trigonometric system is an orthonormal basis in $L^2([0,1])$, however it does not provide an unconditional basis in $L^p([0,1])$ for $p \neq 2$. This reflects a lack of "orthogonality" of the trigonometric system in $L^p([0,1])$. There is a square function that plays the same role as in wavelets, but it involves more than just the absolute value of the Fourier coefficients,

$$Sf(x) = \left(\sum_j |\Delta_j f(x)|^2 \right)^{1/2},$$

where

$$\Delta_j f(x) = \sum_{2^j \le |n| < 2^{j+1}} \hat{f}(n) e^{2\pi i n x}.$$

It is true that $\|f\|_p$ is comparable to $\|S(f)\|_p$ in the sense of (26). We are allowed to change the signs of the Fourier coefficients of f on *the dyadic blocks*. If we denote by $\tilde{f}$ the function reconstructed with the modified coefficients, that is,

$$\tilde{f}(x) = \sum_{j \le 0} \delta_j \Delta_j f(x), \quad \delta_j = \pm 1,$$

then clearly, $S(\tilde{f}) = S(f)$, hence their L^p norms are the same, and are both equivalent to $\|f\|_p$. But there is no guarantee that if we change signs inside the dyadic blocks, this will still be true! In that case, $S(f)$ does not have to coincide with $S(\tilde{f})$. In fact we could "jump" out of $L^p([0,1])$ by performing such innocent operation.

Exercise 39 *Find a simple example to illustrate this property!*

The study of square functions goes under the name *Littlewood-Paley Theory*, and it is a widely used tool in harmonic analysis. We will not say more about it in these lectures.

4 Zooming properties of wavelets, and applications

We want to highlight the "zooming" properties of the Haar system, and how they can be mathematically encoded in the so-called *multiresolution analysis* (MRA). This provides a framework for the construction of most wavelets. We discuss how to implement the wavelet transform via filter banks. We then present, in a very informal manner, competing attributes we would like the wavelets to have, and a by no means exhaustive catalog of wavelets. We then discuss very briefly wavelet packets and two dimensional wavelets used in image processing. We hope to illustrate in the lab how basic compression and denoising of images and signals can be done using wavelet decompositions.

4.1 Multiresolution Analysis

An *orthogonal multiresolution analysis* is a decomposition of $L^2(\mathbb{R})$, into a chain of closed subspaces

$$\cdots \subset \mathbf{V}_{-2} \subset \mathbf{V}_{-1} \subset \mathbf{V}_0 \subset \mathbf{V}_1 \subset \mathbf{V}_2 \subset \cdots \subset L^2(\mathbb{R})$$

such that

1. $\bigcap_{j\in\mathbb{Z}} \mathbf{V}_j = \{0\}$ (trivial intersection),
2. $\bigcup_{j\in\mathbb{Z}} \mathbf{V}_j$ is dense in $L^2(\mathbb{R})$ (density in $L^2(\mathbb{R})$),
3. $f(x) \in \mathbf{V}_j$ if and only if $f(2x) \in \mathbf{V}_{j+1}$ (scaling property),
4. $f(x) \in \mathbf{V}_0$ if and only if $f(x-k) \in \mathbf{V}_0$ for any $k \in \mathbb{Z}$ (shift invariance),
5. There exists a *scaling function* $\phi \in \mathbf{V}_0$ such that its integer translates, $\{\phi(x-k)\}_{k\in\mathbb{Z}}$, form an orthonormal basis of $\mathbf{V}_0$.

For notational convenience we define

$$\phi_{j,k} = 2^{-j/2}\phi(2^{-j}x - k). \tag{28}$$

Exercise 40 *Show that $\{\phi_{j,k}\}_{k\in\mathbb{Z}}$ is an orthonormal basis of $\mathbf{V}_j$.*

Given an L^2- function f, let $P_j f$ be the *orthogonal projection of f into* $\mathbf{V}_j$, i.e.,

$$P_j f = \sum_{k\in\mathbb{Z}} \langle f, \phi_{j,k}\rangle \phi_{j,k}. \tag{29}$$

The function $P_j f$ is an approximation to the original function at scale 2^{-j}. More precisely, it is the *best approximation* in the subspace $\mathbf{V}_j$.

How do we go from the approximation P_jf to the better approximation $P_{j+1}f$? We simply add their difference. Letting

$$Q_jf = P_{j+1}f - P_jf,$$

we clearly have $P_{j+1} = P_j + Q_j$. This defines Q_j to be the projection onto a closed subspace, which we call $\mathbf{W}_j$. The space $\mathbf{W}_j$ is the *orthogonal complement* of $\mathbf{V}_j$ in $\mathbf{V}_{j+1}$. This means that $\mathbf{V}_j \perp \mathbf{W}_j$, and if $f \in \mathbf{V}_{j+1}$, there exist unique $g \in \mathbf{V}_j$ and $h \in \mathbf{W}_j$ such that $f = g + h$. We use the notation

$$\mathbf{V}_{j+1} = \mathbf{V}_j \oplus \mathbf{W}_j,$$

so that

$$L^2(\mathbb{R}) = \bigoplus_{j\in\mathbb{Z}} \mathbf{W}_j. \tag{30}$$

One can show, see [Mall89], that the scaling function ϕ, determines the *wavelet* ψ, such that $\{\psi(x-k)\}_{k\in\mathbb{Z}}$ is an orthonormal basis of $\mathbf{W}_0$. Since $\mathbf{W}_j$ is a dilation of $\mathbf{W}_0$, we can define

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

and the family $\{\psi_{j,k}\}_{k\in\mathbb{Z}}$ forms an orthonormal basis of $\mathbf{W}_j$.

A calculation shows,

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

All together, these functions will provide a wavelet basis.

Theorem 7 (Mallat) *Given an MRA with scaling function ϕ, there is a wavelet $\psi \in L^2(\mathbb{R})$ such that for each j, the family $\{\psi_{j,k}\}_{k\in\mathbb{Z}}$ is an orthonormal basis of $\mathbf{W}_j$. Hence the family $\{\psi_{j,k}\}_{j,k\in\mathbb{Z}}$ is an orthonormal basis of $L^2(\mathbb{R})$.*

Notice that on the Fourier side,

$$(\psi_{j,0})^\wedge(\xi) = 2^{-j/2}\hat{\psi}(2^{-j}\xi).$$

The Heisenberg boxes change with the scale parameter j, and remain constant along translations k. For $j = 0$ they are essentially squares of area 1, for other j the dimensions are $2^j \times 2^{-j}$. The wavelet transform divides the phase plane differently than either the Fourier or Gabor bases. The wavelet phase plane is given in Figure 1.

Example 13 *The characteristic function $\phi(t) = \chi_{[0,1]}(t)$ is the scaling function of an orthogonal MRA. The subspace $\mathbf{V}_j$ corresponds to step functions with steps on the intervals $[k2^{-j}, (k+1)2^{-j}]$, and they have all the properties listed. The Haar function is the wavelet ψ in Mallat's theorem. We will say more about this example in the next section.*

Are there other MRAs? Yes, there are. In the case of the Shannon wavelet, the scaling function satisfies $\hat{\phi}(\xi) = \chi_{[-1/2,1/2]}(\xi)$.

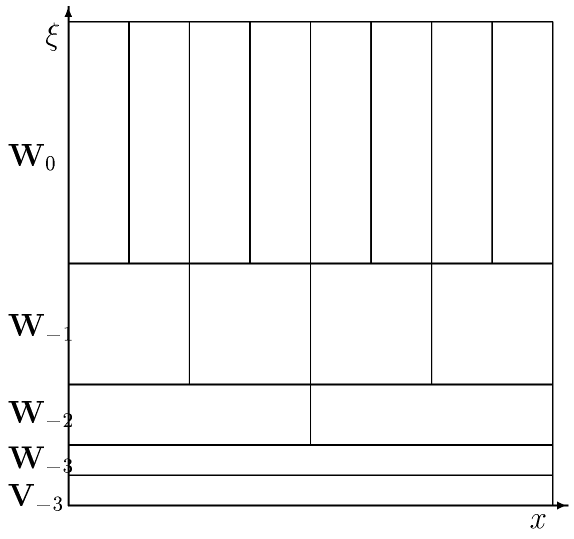

Figure 1: The wavelet phase plane.

Exercise 41 *Describe the MRA generated by the Shannon scaling function.*

While there are wavelets that do not come from an MRA, these are rare. If the wavelet has compact support then it does come from an MRA. Finally, the conditions in the definition of the MRA are not independent but for most applications are sufficient. For a full account of all these issues and more consult the book by Hernández and Weiss [HW, Chapter 2].

4.1.1 Haar MRA

Before considering how to construct scaling functions and wavelets, we will consider a simple example that predates the main development of wavelets, the Haar multiresolution analysis.

Let the scaling function be

$$\phi(x) = \begin{cases} 1 & \text{for } \ 0 \le x < 1 \\ 0 & \text{elsewhere} \end{cases} .$$

Then $\mathbf{V}_0 = \mathrm{span}(\{\phi(x-k)\}_{k\in\mathbb{Z}})$ consists of piecewise constant functions with jumps only at integers. The wavelet is

$$\psi(x) = h(x) = \begin{cases} -1 & \text{for } \ 0 \le x < 1/2 \\ 1 & \text{for } \ 1/2 \le x < 1 \\ 0 & \text{elsewhere} \end{cases} .$$

The subspace $\mathbf{W}_0 = \mathrm{span}(\{\psi(x-k)\}_{k\in\mathbb{Z}})$ are piecewise constant functions with jumps only at half-integers, and average 0 between integers.

We next consider an example of how to decompose a function into its projections onto the subspaces. In practice we select a coarsest scale $\mathbf{V}_{-n}$ and finest scale $\mathbf{V}_0$, truncate the chain to

$$\mathbf{V}_{-n} \subset \cdots \subset \mathbf{V}_{-2} \subset \mathbf{V}_{-1} \subset \mathbf{V}_0$$

and obtain

$$\mathbf{V}_0 = \mathbf{V}_{-n} \bigoplus_{j=1}^{n} \mathbf{W}_{-j}. \tag{31}$$

We will go through the decomposition process in the text using vectors, but it is more enlightening to look at the graphical version in Figure 2.

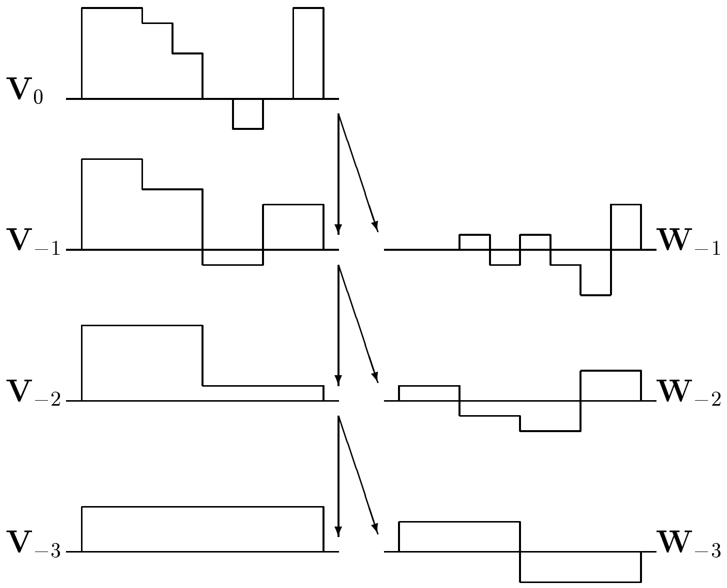

Figure 2: A wavelet decomposition. $\mathbf{V}_0 = \mathbf{V}_{-3} \oplus \mathbf{W}_{-3} \oplus \mathbf{W}_{-2} \oplus \mathbf{W}_{-1}$.

We begin with a vector of $8 = 2^3$ "samples" of a function, which we assume to be the average value of the function on 8 intervals of length 1, so that our function is supported on the interval $[0, 8]$. We will use the vector

$$v_0 = [6, 6, 5, 3, 0, -2, 0, 6]$$

to represent our function in $\mathbf{V}_0$. To construct the projection onto $\mathbf{V}_{-1}$ we average pairs of values, to obtain

$$v_{-1} = [6, 6, 4, 4, -1, -1, 3, 3].$$

The difference is in $\mathbf{W}_{-1}$, so we have

$$w_{-1} = [0, 0, 1, -1, 1, -1, -3, 3].$$

By repeating this process, we obtain

$$
\begin{aligned}
v_{-2} &= [5,5,5,5,1,1,1,1], \\
w_{-2} &= [1,1,-1,-1,-2,-2,2,2], \\
v_{-3} &= [3,3,3,3,3,3,3,3], \quad \text{and} \\
w_{-3} &= [2,2,2,2,-2,-2,-2,-2].
\end{aligned}
$$

For the example in Figure 2, the scaling function subspaces are shown in Figure 3 and the wavelet subspaces are shown in Figure 4. To compute the coefficients of the expansion

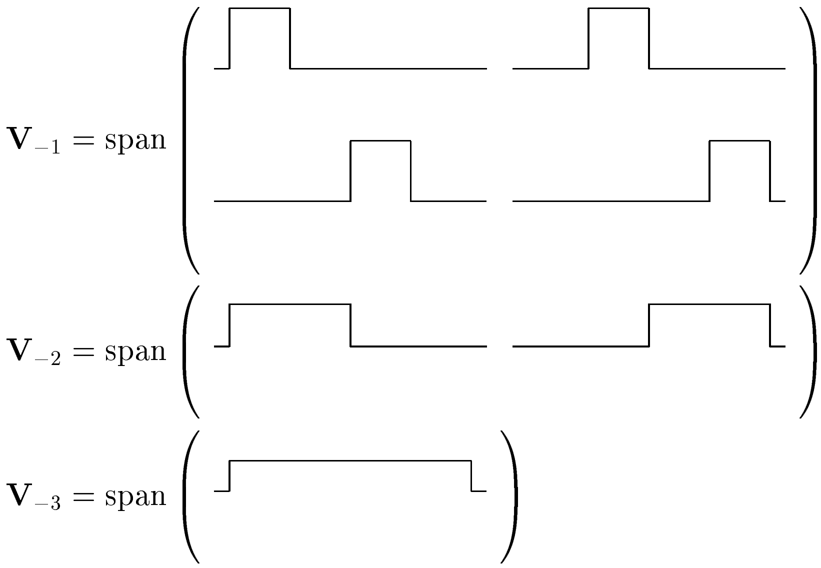

Figure 3: The scaling function subspaces used in Figure 2.

(29), we need to compute the inner product $\langle f, \phi_{j,k} \rangle$ for the function (28). In terms of our vectors, we have, for example

$$\langle f, \phi_{0,3} \rangle = \langle [6,6,5,3,0,-2,0,6], [0,0,0,1,0,0,0,0] \rangle = 3$$

and

$$\langle f, \phi_{1,1} \rangle = \langle [6,6,5,3,0,-2,0,6], [0,0,1/\sqrt{2},1/\sqrt{2},0,0,0,0] \rangle = 8/\sqrt{2}$$

The scaling function ϕ satisfies the *two-scale recurrence equation*,

$$\phi(t) = \phi(2t) + \phi(2t-1),$$

which tells us $\phi_{j,k} = (\phi_{j-1,2k} + \phi_{j-1,2k+1})/\sqrt{2}$, and thus

$$\langle f, \phi_{j,k} \rangle = \frac{1}{\sqrt{2}}(\langle f, \phi_{j-1,2k} \rangle + \langle f, \phi_{j-1,2k+1} \rangle).$$

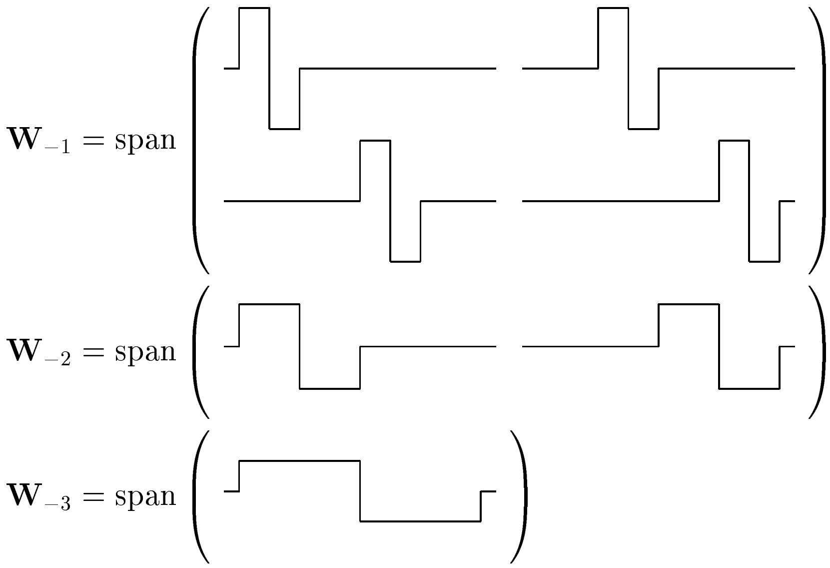

Figure 4: The wavelet subspaces used in Figure 2.

Thus we can also compute

$$\langle f, \phi_{1,1} \rangle = \frac{1}{\sqrt{2}}(5+3).$$

The coefficients $\langle f, \phi_{j,k} \rangle$ for j fixed are called the *averages* of f at scale j, and commonly denoted $a_{j,k}$.

Similarly, the wavelet satisfies the *two-scale difference equation*,

$$\psi(t) = \phi(2t) - \phi(2t-1),$$

and thus we can recursively compute

$$\langle f, \psi_{j,k} \rangle = \frac{1}{\sqrt{2}}(\langle f, \phi_{j-1,2k} \rangle - \langle f, \phi_{j-1,2k+1} \rangle).$$

The coefficients $\langle f, \psi_{j,k} \rangle$ for j fixed are called the *differences* of f at scale j, and commonly denoted $d_{j,k}$. Evaluating the whole set of Haar coefficients $d_{j,k}$ and averages $a_{j,k}$ requires $2(N-1)$ additions and $2N$ multiplications. The discrete wavelet transform can be performed using a similar *cascade algorithm* with complexity N, where N is the number of data points.

We can view the averages at resolution j like successive approximations to the original signal, these are the orthogonal projections $P_j f$ onto the approximation spaces $\mathbf{V}_j$ (expectation operators defined in Section 3.3.3). The *details*, necessary to move from level j to the next level $(j+1)$, are encoded in the Haar coefficients at level j, more precisely in the orthogonal projections $Q_j f$ onto the detail subspaces $\mathbf{W}_j$ (difference operators defined in Section 3.3.3). Starting at a low resolution level, we can obtain better and better resolution by adding the details at the subsequent levels. As $j \to \infty$, the resolution is increased. The steps get smaller (length 2^{-j}), and the approximation converges to $f \in L^2(\mathbb{R})$ in L^2-norm (this is the content of (22) in Theorem 6), and a.e. (Lebesgue Differentiation Theorem). Clearly the subspaces are nested, that is, $\mathbf{V}_j \subset \mathbf{V}_{j+1}$, and their intersection is the trivial subspace containing just the zero function (this is (21) in Theorem 6). Lo and behold, we have shown that the Haar scaling function generates an MRA.

4.2 Filter Banks

Here we want to give a glimpse of how we would implement the wavelet transform once an MRA is at our disposal, in a similar way to what we just did for the Haar functions.

The claim is that all one needs for computations is the so called *filter coefficients* and not the scaling and wavelet functions. These are finite sequences of numbers if and only if the scaling function is *compactly supported*, like it is the case of the Haar scaling function.

Given an orthogonal MRA with scaling function ϕ, then $\phi \in \mathbf{V}_0 \subset \mathbf{V}_1$, and the functions $\phi_{1,k}(t) = \sqrt{2}\phi(2t-k)$, for $k \in \mathbb{Z}$, form an orthonormal basis of $\mathbf{V}_1$. This means that the following *scaling equation* holds, for some coefficients $\{h_k\}$ such that $\sum_k |h_k|^2 < \infty$,

$$\phi(t) = \sum_{k\in\mathbb{Z}} h_k \phi_{1,k}(t) = \sqrt{2}\sum_{k\in\mathbb{Z}} h_k \phi(2t-k).$$

In general this sum is not finite, but whenever it is, the scaling function ϕ has *compact support*, and so will the wavelet. In the case of the Haar MRA, we have $h_0 = h_1 = 1/\sqrt{2}$, and all other coefficients vanish. The sequence $H = \{h_k\}$ is the so called *low-pass filter*. We will assume that the low-pass filter has length L (it turns out that such filters will always have even lengths $L = 2M$). The *refinement mask* is given by $H(\xi) = \frac{1}{\sqrt{2}}\sum h_k e^{2\pi i k\xi}$, which can be viewed as a periodic function of the frequency variable ξ with period one, whose Fourier coefficients are $\hat{H}(n) = h_n/\sqrt{2}$.

Not all wavelets have compact support. However for applications it is a most desirable property since it corresponds to the FIR filter case.

The existence of a solution of the scaling equation can be expressed in the language of fixed point theory. Given a low-pass filter H, define a transformation $T\phi(t) = \sqrt{2}\sum_k h_k\phi(2t-k)$, does it have a fixed point? If yes, then the fixed point is a solution to the scaling equation.

Exercise 42 *Check that on Fourier side the scaling equation becomes,*

$$\hat{\phi}(\xi) = H(\xi/2)\hat{\phi}(\xi/2).$$

We can iterate this formula to obtain,

$$\hat{\phi}(\xi) = \left(\Pi_{j=0}^{N} H(\xi/2^j)\right)\hat{\phi}(\xi/2^N).$$

Provided $\hat{\phi}$ is continuous at $\xi = 0$, $\hat{\phi}(0) \neq 0$, and the infinite product converges, then a solution to the scaling equation will exist. It turns out that to obtain orthonormality of the set $\{\phi_{0,k}\}_{k\in\mathbb{Z}}$ we must have $|\hat{\phi}(0)| = 1$, and one usually normalizes to $\hat{\phi}(0) = \int \phi(t)dt = 1$, which happens to be useful in numerical implementations of the wavelet transform.

By now the conditions on the filter H that will guarantee the existence of a solution ϕ to the scaling equation are well understood (consult [HW] for more details). For example, necessarily $H(0) = 1$, or equivalently $\sum_{k=0}^{L-1} h_k = \sqrt{2}$. After some manipulations, one can conclude that the orthonormality of the integer shifts of the scaling function imply

$$|H(\xi)|^2 + |H(\xi + 1/2)|^2 = 1. \tag{32}$$

This would be recognized in the engineering community as a *quadrature mirror filter* (QMF) condition, necessary for exact reconstruction for a pair of filters.

Notice that the wavelet ψ we are seeking is an element of $\mathbf{W}_0 \subset \mathbf{V}_1$. Therefore it will also be a superposition of the basis $\{\phi_{1,k}\}_{k\in\mathbb{Z}}$ of $\mathbf{V}_1$. Hence there are coefficients $\{g_k\}$ such that,

$$\psi(t) = \sum_{k\in\mathbb{Z}} g_k \phi_{1,k}(t).$$

If G is a QMF, then for each scale j, the wavelets $\{\psi_{j,k}\}_{k\in\mathbb{Z}}$ form an orthonormal basis for $\mathbf{W}_j$. The orthogonality between $\mathbf{W}_0$ and $\mathbf{V}_0$ implies that for all ξ,

$$H(\xi)\overline{G(\xi)} + H(\xi + 1/2)\overline{G(\xi + 1/2)} = 0, \tag{33}$$

Define the high-pass filter $G = \{g_k\}$ by,

$$g_k = (-1)^{k-1}\overline{h_{1-k}}, \quad G(\xi) = \frac{1}{\sqrt{2}}\sum_{k=0}^{L-1} g_k e^{2\pi i k\xi}. \tag{34}$$

Exercise 43 *Check that $G(\xi) = e^{2\pi i\xi}\overline{H(\xi + 1/2)}$, and that if H satisfies a QMF condition, so does G. Check that on Fourier side,*

$$\hat{\psi}(\xi) = G(\xi/2)\hat{\phi}(\xi/2),$$

and that ψ is orthogonal to $\phi_{0,k}$ for all $k \in \mathbb{Z}$.

Check that (33) *is automatic with this choice of G.*

Example 14 (Haar revisited) *The characteristic function of the unit interval $\chi_{[0,1]}$ generates an orthogonal MRA (Haar MRA). The non-zero low-pass filter coefficients are $h_0 = h_1 = 1/\sqrt{2}$; hence the non-zero high-pass coefficients are $g_0 = -1/\sqrt{2}$ and $g_1 = 1/\sqrt{2}$. Therefore the Haar wavelet is $\psi(t) = \phi(2t-1) - \phi(2t)$. The refinement masks are $H(\xi) = \frac{1+e^{2\pi i\xi}}{2}$, $G(z) = \frac{e^{-2\pi i\xi}-1}{2}$.*

Example 15 (Shannon revisited) *This time $\hat{\phi}(\xi) = \chi_{[-1/2,1/2]}(\xi)$ generates an orthogonal MRA. One can deduce from Exercise* 42 *that $H(\xi) = \chi_{\{1/4<|\xi|\leq 1/2\}}(\xi)$. Hence by Exercise* 43*,*

$$G(\xi) = e^{2\pi i\xi}H(\xi + 1/2) = e^{2\pi i\xi}\chi_{[-1/4,1/4]}(\xi),$$

(remember we are viewing $H(\xi)$ and $G(\xi)$ periodic functions on the unit interval), and

$$\hat{\psi}(\xi) = e^{\pi i\xi}\chi_{\{1/2<|\xi|\leq 1\}}(\xi).$$

Example 16 (Daubechies wavelets) *For each integer $N \geq 1$ there is an orthogonal MRA that generates a compactly and minimally supported wavelet (length of the support $2N$), and the filters have $2N$-taps. They are denoted in Matlab by dbN. The db1 corresponds to the Haar wavelet. The coefficients corresponding to db2 are:*

$$h_0 = \frac{1+\sqrt{3}}{4\sqrt{2}}, \qquad h_1 = \frac{3+\sqrt{3}}{4\sqrt{2}}, \qquad h_2 = \frac{3-\sqrt{3}}{4\sqrt{2}}, \qquad h_3 = \frac{1-\sqrt{3}}{4\sqrt{2}}.$$

Exercise 44 *Check that db2 filter is a QMF, and that $h_0 + h_1 + h_2 + h_3 = \sqrt{2}$.*

It turns out that for finite filters H, the conditions (32), and $\sum_k h_k = \sqrt{2}$, are sufficient to guarantee the existence of a solution ϕ to the scaling equation. For infinite filters an extra decay assumption is necessary. However, it is not sufficient to guarantee the orthonormality of the integer shifts of ϕ. But, if for example $\inf_{1/4\leq|\xi|\leq 3/4} |H(\xi)| > 0$ is also true, then $\{\phi_{0,k}\}$ is an orthonormal set in $L^2(\mathbb{R})$. See for example [Fr, Ch. 5] for more details, or more advanced textbooks.

A *cascade algorithm*, similar to the one described for the Haar basis can be implemented to provide a fast wavelet transform. Given the coefficients $\{a_{J,k}\}_{k=0,1,\ldots,N-1}$, $N = 2^J$ "scaled samples" of the function f defined on the interval $[0,1]$ and extended periodically on the line. Then the coefficients $a_{j,k}$ and $d_{j,k}$ for scales $j < J$ can be calculated in order LN operations, where L is the length of the filter, via,

$$a_{j,k} = \sum_{n=0}^{N2^{-j}-1} \overline{h_n}\, a_{j+1,n+2k} = \tilde{H} * a_{j+1}(2k),$$

$$d_{j,k} = \sum_{n=0}^{N2^{-j}-1} \overline{g_n}\, a_{j+1,n+2k} = \tilde{G} * a_{j+1}(2k),$$

where $a_j = \{a_{j,k}\}$ and $d_j = \{d_{j,k}\}$ are viewed as periodic sequences with period $N2^{-j} > L$ (effectively in each sum there are only L non-zero terms). And $\tilde{H} = \{\tilde{h}_k = \overline{h_{-k}}\}$ is the conjugate flip of the filter H, similarly for $\tilde{G}$. We can obtain the approximation and detail coefficients at a rougher scale $(j-1)$ by convolving (circular convolution) the approximation coefficients at scale j with the low and high-pass filters $\tilde{H}$ and $\tilde{G}$, and *downsampling* by a factor of 2. More precisely the *downsampling operator* takes an N-vector and maps it into a vector half as long by discarding the odd entries,

$$Ds(n) = s(2n);$$

it is denoted by the symbol $\downarrow 2$. In electrical engineering terms this is the *analysis phase* of a *subband filtering scheme*, which can be represented schematically by:

$$\begin{tikzcd} & \boxed{*\tilde{H}} \arrow[r] & \boxed{\downarrow 2} \arrow[r] & a_j \\ a_{j+1} \arrow[ur] \arrow[dr] & & & \\ & \boxed{*\tilde{G}} \arrow[r] & \boxed{\downarrow 2} \arrow[r] & d_j \end{tikzcd}$$

If the function f is not periodic, then such periodization will create artifacts at the boundaries. In that case, it is better to use wavelets adapted to the interval, which we will not discuss further in these lectures.

Computing the coefficients can be represented by the following tree or *pyramid scheme*,

$$\begin{array}{ccccccccc} a_J & \rightarrow & a_{J-1} & \rightarrow & a_{J-2} & \rightarrow & a_{J-3} & \cdots \\ & \searrow & & \searrow & & \searrow & & \\ & & d_{J-1} & & d_{J-2} & & d_{J-3} & \cdots \end{array}$$

The reconstruction of the "samples" at level j from the samples and details at the previous level, is also an order N algorithm,

$$a_{j+1,2k} = \sum_{n=1}^{L/2=M} h_{2k} a_{j,n-k} + \sum_{n=1}^{L/2=M} g_{2k} d_{j,n-k},$$

$$a_{j+1,2k-1} = \sum_{n=1}^{L/2=M} h_{2k-1} a_{j,n-k} + \sum_{n=1}^{L/2=M} g_{2k-1} d_{j,n-k}.$$

Which can also be rewritten in terms of circular convolutions:

$$a_{j+1,m} = \sum_n \left[h_{m-2n} a_{j,n} + g_{m-2n} d_{j,n} \right] = H * U a_j(m) + G * U d_j(m),$$

where here we first upsample the approximation and detail coefficients to restore the right dimensions, then convolve with the filters, and finally add the outcomes. More precisely the *upsampling operator* takes an N-vector and maps it to a vector twice as long, by intertwining zeros,

$$Us(n) = \begin{cases} s(n/2) & \text{if } n \text{ is even} \\ 0 & \text{if } n \text{ is odd} \end{cases};$$

it is denoted by the symbol $\uparrow 2$. This corresponds to the synthesis phase of the *subband filtering scheme*, which can be represented schematically by:

$$\begin{array}{ccccccc} a_j \rightarrow \boxed{\uparrow 2} \rightarrow \boxed{*H} & & & & \\ & \searrow & & & \\ & & \oplus \rightarrow a_{j+1} & & \\ & \nearrow & & & \\ d_j \rightarrow \boxed{\uparrow 2} \rightarrow \boxed{*G} & & & & \end{array}$$

Exercise 45 *Compute the Fourier transform for both the up and downsampling operators in finite dimensional space.*

The reconstruction can also be illustrated with a pyramid scheme,

$$\begin{array}{ccccccccc} a_n & \to & a_{n+1} & \to & a_{n+2} & \to & a_{n+3} & \cdots \\ & \nearrow & & \nearrow & & \nearrow & & \\ d_n & & d_{n+1} & & d_{n+2} & & \cdots & \end{array}$$

Note that once the low-pass filter H is chosen, everything else, high-pass filter G, scaling function ϕ, and wavelet ψ are completely determined, and so is the MRA. In practice one never computes the values of ϕ and ψ. All the manipulations are performed with the filters G and H, even if they involve calculating quantities associated to ϕ or ψ, like moments or derivatives. However we might want to produce pictures of the wavelet and scaling functions from the filter coefficients.

Example 17 *The cascade algorithm can be used to produce very good approximations for both ψ and ϕ, and this is how pictures of the wavelets and the scaling functions are obtained. For the scaling function ϕ, it suffices to observe that $a_{1,k} = \langle\phi, \phi_{1,k}\rangle = h_k$ and $d_{j,k} = \langle\phi, \psi_{j,k}\rangle = 0$ for all $j \geq 1$ (the first because of the scaling equation, the second because $\mathbf{V}_0 \subset \mathbf{V}_j \perp \mathbf{W}_j$ for all $j \geq 1$), that is what we need to initialize and iterate as many times as we wish (say n iterations) the synthesis phase of the filter bank,*

$$H \to \boxed{\uparrow 2} \to \boxed{\;*H\;} \to \boxed{\uparrow 2} \to \boxed{\;*H\;} \to \cdots \to \boxed{\uparrow 2} \to \boxed{\;*H\;} \to \{\langle\phi, \phi_{n+1,k}\rangle\}.$$

The output after n-iterations is the set of approximation coefficients at scale $j = n+1$. After multiplying by a scaling factor one can make precise the statement that

$$\psi(k2^{-j}) \sim 2^{-j/2}\langle\phi, \phi_{j,k}\rangle.$$

A graph can now be plotted (at least for real-valued filters).

This is done similarly for the wavelet ψ. Notice that this time $a_{1,k} = \langle\psi, \phi_{1,k}\rangle = g_k$ and $d_{j,k} = \langle\psi, \psi_{j,k}\rangle = 0$ for all $j \geq 1$. The cascade algorithm now will produce the approximation coefficients at scale j after $n = j-1$ iterations,

$$G \to \boxed{\uparrow 2} \to \boxed{\;*H\;} \to \boxed{\uparrow 2} \to \boxed{\;*H\;} \to \cdots \to \boxed{\uparrow 2} \to \boxed{\;*H\;} \to \{\langle\psi, \phi_{j,k}\rangle\}.$$

All together we are facing a *perfect reconstruction filter bank*,

$$\begin{array}{ccccc} & & \boxed{*\tilde{H}} \to \boxed{\downarrow 2} \to a_{j-1} \to \boxed{\uparrow 2} \to \boxed{*H} & & \\ & \nearrow & & \searrow & \\ a_j \to & & & & \oplus \to a_j \\ & \searrow & & \nearrow & \\ & & \boxed{*\tilde{G}} \to \boxed{\downarrow 2} \to d_{j-1} \to \boxed{\uparrow 2} \to \boxed{*G} & & \end{array}$$

which does not always involve the same filters in the analysis and synthesis phase. When it does, as in our case, we have an *orthogonal filter bank*. One could obtain perfect reconstruction in more general cases, if we replace $\tilde{H}$ by a filter $\tilde{H}^*$, and $\tilde{G}$ by $\tilde{G}^*$, then we would have a *biorthogonal filter bank*, with dual filters H, H^*, G, G^*.

In the case of a biorthogonal filter bank, there will be an associated *biorthogonal MRA* with *dual scaling functions* ϕ, ϕ^*, and under certain conditions they will generate a *biorthogonal wavelet basis* $\{\psi_{j,k}\}$ and *dual basis* $\{\psi^*_{j,k}\}$, so that

$$f(x) = \sum_{j,k} \langle f, \psi^*_{j,k}\rangle \psi_{j,k} = \sum_{j,k} \langle f, \psi_{j,k}\rangle \psi^*_{j,k}.$$

The following substitute for Plancherel holds: (*Riesz basis property*) there exist $A, B > 0$ such that for all $f \in L^2(\mathbb{R})$,

$$\sum_{j,k} |\langle f, \psi^*_{j,k}\rangle|^2 \sim \|f\|_2^2 \sim \sum_{j,k} |\langle f, \psi_{j,k}\rangle|^2,$$

where $A \sim B$ means that there exit constants $c, C > 0$ such that $cA \leq B \leq CA$. The relative size of the similarity constants c, C becomes important for numerical calculations, it is related to the *condition number* of a matrix.

Exercise 46 *Consider two linearly independent vectors in* $\mathbb{R}^2$, $\vec{v}_1 = (x_1, y_1)$, $\vec{v}_2 = (x_2, y_2)$. *They form a basis of* $\mathbb{R}^2$. *Find "dual vectors"* $\vec{v}_1^*, \vec{v}_2^* \in \mathbb{R}^2$, *such that, given any vector* $\vec{u} \in \mathbb{R}^2$,

$$\vec{u} = \langle \vec{u}, \vec{v}_1^*\rangle \vec{v}_1 + \langle \vec{u}, \vec{v}_2^*\rangle \vec{v}_2.$$

Show that $\vec{u} = \langle \vec{u}, \vec{v}_1\rangle \vec{v}_1^* + \langle \vec{u}, \vec{v}_2\rangle \vec{v}_2^*$. *Find positive numbers* c, C, *such that,*

$$c\sum_{j=1}^{2} |\langle \vec{u}, \vec{v}_j^*\rangle|^2 \leq |\vec{u}|^2 \leq C\sum_{j=1}^{2} |\langle \vec{u}, \vec{v}_j^*\rangle|^2.$$

Draw a picture explaining orthogonality properties among these vectors! The numbers c, C *should be related to the angle between the vectors* $\vec{v}_1$ *and* $\vec{v}_1$.

A general filter bank is any sequence of convolutions and other simple operations like up and down sampling. The study of such banks is an entire subject in engineering called *multirate signal analysis*, or *subband coding*. The term *filter* is used to denote a convolution operator because such operator can cut out various frequencies if the corresponding Fourier multiplier vanishes (or is very small) at those frequencies. Consult Strang and Nguyen's textbook [SN] for more information. Filter banks can be implemented quickly because of the Fast Fourier Transform. Remember that convolution becomes multiplication on the Fourier side, that is a diagonal matrix (need only N products, whereas ordinary matrix multiplication requires N^2 products), and to go back and forth from Fourier to time domain, we can use the Fast Fourier Transform, so the total number of operations will be of the order $N \log N$.

There exist commercial and free software dealing with these applications. The main commercial one is the *Matlab Wavelet Toolbox* which we will use in the lab. *Wavelab* is a Stanford based free software which is Matlab-based. *Lastwave* is a toolbox with subroutines written in C, with "a friendly shell and display environment" according to Mallat, elaborated at Ecole Polytechnique. There is much more in the internet, and it suffices to make a search for wavelets to see the enormous amount of information, codes, etc. available online.

4.3 Design Features

Most of the applications of wavelets exploit their ability to approximate functions as efficiently as possible, that is with as few coefficients as possible. For different applications one wishes the wavelets to have various properties. Some of them are competing against each other, so it is up to the user to decide which one is most efficient for her problem. The most popular conditions are orthogonality, compact support, vanishing moments, symmetry and smoothness.

Orthogonality: Orthogonality allows for straightforward calculation of the coefficients (via inner products with the basis elements). It guarantees that energy is preserved. However sometimes orthogonality can be substituted by *biorthogonality*. In this case, there is an auxiliary set of dual functions that is used to compute the coefficients by taking inner products; also the energy is almost preserved.

Compact support: We have already stressed that compact support is important for numerical purposes (implementation of the FIR). In terms of detecting point singularities, it is clear that if the signal f has a singularity at t_0 then if t_0 is inside the support of $\psi_{j,n}$, the corresponding coefficient could be large. If ψ has support of length l, then at each scale j there will be l wavelets interacting with the singularity (that is their support contains t_0). The shorter the support the fewer wavelets interacting with the singularity.

We have already mentioned that compact support of the scaling function coincides with FIR. Moreover if the low-pass filter is supported on $[N_1, N_2]$, so is ϕ, and it is not hard to see that ψ will have support of the same length $(N_2 - N_1)$ but centered at $1/2$.

Smoothness: The regularity of the wavelet has effect on the error introduced by thresholding or quantizing the wavelet coefficients. Suppose an error ϵ is added to the coefficient $\langle f, \psi_{j,k}\rangle$, then we will add an error of the form $\epsilon\psi_{j,k}$ to the reconstruction. Smooth errors are often less *visible* or *audible*. Often better quality images are obtained when the wavelets are smooth. However, the smoother the wavelet, the longer the support.

There is no orthogonal wavelet that is C^∞ and has exponential decay. Therefore there is no hope of finding an orthogonal wavelet that is C^∞ and has compact support.

Vanishing moments: A function ψ has p vanishing moments if for all $k = 0, 1, \ldots, p-1$,

$$\int_{\mathbb{R}} t^k \psi(t)dt = 0.$$

This automatically implies that ψ is orthogonal to polynomials of degree $p-1$. Smooth functions f will have small fine scale wavelet coefficients. More precisely, if the function to be analysed is k-regular, it can be approximated well by a Taylor polynomial of degree k. If $k < p$, then the wavelets are orthogonal to that Taylor polynomial, and the coefficients are small. If ψ has p vanishing moments, then the polynomials of degree $p-1$ are reproduced by the scaling functions, this is often referred as the *approximation order* of the MRA.

The constraints imposed on orthogonal wavelets imply that if ψ has p moments then its support is at least of length $2p-1$. Daubechies wavelets have minimum support length for a given number of vanishing moments. So there is a trade-off between length of the support and vanishing moments. If the function has few singularities and is smooth between singularities, then we might as well take advantage of the vanishing moments. If there are many singularities, we might prefer to use wavelets with shorter supports.

Symmetry: It is impossible to construct compactly supported symmetric orthogonal wavelets except for Haar. However symmetry is often useful for image and signal analysis. It can be obtained at the expense of one of the other properties. If we give up orthogonality, then there are compactly supported, smooth and symmetric biorthogonal wavelets. If we use *multiwavelets*[8], we can construct them to be orthogonal, smooth, compactly supported and symmetric. Some wavelets have been designed to be nearly symmetric (Daubechies symmlets, for example).

4.4 A catalog of wavelets

Here we will list the main wavelets and indicate their properties.

Haar wavelet: Perfectly localized in time, not so good in frequency, discontinuous. Symmetric. Shortest possible support, only one vanishing moment, hence not well adapted to approximating smooth functions.

Shannon wavelet: This wavelet does not have compact support, however it is C^∞. It is band-limited, but its Fourier transform is discontinuous, hence $\psi(t)$ decays like $1/|t|$ at infinity. $\hat{\psi}(\xi)$ is zero in a neighborhood of $\xi = 0$, hence all its derivatives are zero at $\xi = 0$, hence ψ has an infinite number of vanishing moments.

Meyer wavelet: This is a symmetric band-limited function whose Fourier transform is smooth, hence $\psi(t)$ has faster decay at infinity. The scaling function is also band-limited. Hence both ψ and ϕ are C^∞. ψ has an infinite number of vanishing moments. (This wavelet was found by Strömberg in 1983, and it went unnoticed for several years).

Battle-Lemarié spline wavelets: Polynomial splines of degree m. ψ has $m+1$ vanishing moments. They don't have compact support, but they have exponential decay. Since they are polynomial splines of degree m, they are $m-1$ times continuously differentiable. For m odd, ψ is symmetric around $1/2$. For m even it is antisymmetric around $1/2$. The linear spline wavelet is the Franklyn wavelet.

Daubechies compactly supported wavelets: They have compact support of minimal length for any given number of vanishing moments. More precisely, if ψ has p vanishing moments, then the filters have length $2p$ (or $2p$ *taps*). For large p, ϕ and ψ are uniformly Lipschitz α of the order $\alpha \sim 0.2p$. They are asymmetric. When $p = 1$ we recover the Haar wavelet.

[8]In this case one has more than one scaling function and more than one wavelet function, whose dilates and translates will provide basis in the MRA.

Daubechies symmlets: p vanishing moments, minimum support of length $2p$, as symmetric as possible.

Coiflets: ψ has p vanishing moments, minimum support, and ϕ has $p-1$ moments vanishing (from the second to the pth moment, never the first since $\int \phi = 1$. This extra property requires enlarging the support of ψ to length $(3p-1)$. This time if we approximate a regular function f by a Taylor polynomial, the approximation coefficients will satisfy,

$$2^{J/2}\langle f, \phi_{J,k}\rangle \sim f(2^J k) + O(2^{-(k+1)J}).$$

Hence at fine scale J, the approximation coefficients are close to the signal samples.

The coiflets were constructed by Daubechies after Coifman requested them for the purpose of applications to almost diagonalization of singular integral operators, we will come back to this in the last lecture.

Mexican hat: Has a closed formula involving second derivatives of the Gaussian:

$$\psi(t) = C(1-t^2)e^{t^2/2},$$

where the constant is chosen to normalize it in L^2. It does not come from an MRA, and it is not orthogonal. It is appropriate for continuous wavelet transform. It has exponential decay but not compact support. According to Daubechies, this function is popular in "vision analysis".

Morlet wavelet: It has a closed formula,

$$\psi(t) = Ce^{-t^2/2}\cos(5t).$$

It does not come from an MRA, and it is not orthogonal. It is appropriate for continuous wavelet transform. It has exponential decay but not compact support.

Spline biorthogonal wavelets: These are compactly supported. There are two positive integer parameters N, N^*. N^* determines the scaling function ϕ^*, it is a spline of order $[N^*/2]$. The other scaling function and both wavelets depend on both parameters. ψ^* is a compactly supported piecewise polynomial of order N^*-1, which is C^{N^*-2} at the knots, and increasing support with N. ψ has also support increasing with N and vanishing moments as well. Their regularity can differ notably. The filter coefficients are dyadic rationals, which makes them very attractive for numerical purposes. The functions ψ are known explicitly. The dual filters are very unequal in length, which could be a nuisance when performing image analysis for example.

All the previous wavelets are encoded in Matlab. One can review them and their properties on line.

4.5 Wavelet packets

To perform the wavelet transform we iterate at the level of the low-pass filter (approximation). In principle it is an arbitrary choice, one could iterate at the high-pass filter level, or any desirable combination. The full dyadic tree will provide an overabundance of information, and it corresponds to the *wavelet packets*. Each finite wavelet packet has the information for reconstruction in many different bases, including the wavelet basis. Fast algorithms to search the *best basis* exist. The *Haar packet* includes the Haar basis and the *Walsh basis*. Wickerhauser's book is a good source of information for this topic [Wick].

Denote the spaces by $\mathbf{W}_{j,n}$, where j is the scale as before, and n determines the "frequency".

$$
\begin{array}{ccccc}
 & & & \nearrow & \mathbf{W}_{-2,3} \\
 & & \mathbf{W}_{-1,1} & & \\
 & \nearrow & & \searrow & \mathbf{W}_{-2,2} \\
\mathbf{W}_{0,0} & & & & \\
 & \searrow & & \nearrow & \mathbf{W}_{-2,1} \\
 & & \mathbf{W}_{-1,0} & & \\
 & & & \searrow & \mathbf{W}_{-2,0}
\end{array}
\qquad\qquad
\begin{array}{ccccc}
 & & \mathbf{W}_{-1,1} & & \\
 & \nearrow & & & \\
\mathbf{W}_{0,0} & & & & \\
 & \searrow & & \nearrow & \mathbf{W}_{-2,1} \\
 & & \mathbf{W}_{-1,0} & & \\
 & & & \searrow & \mathbf{W}_{-2,0}
\end{array}
$$

1. Haar packet binary tree - 2 levels. Leaves determine Walsh basis.

2. Connected binary subtree. Leaves determine Haar basis.

Notice that $\mathbf{W}_{0,0} = \mathbf{V}_0$, more generally, $\mathbf{W}_{j,0} = \mathbf{V}_j$, and $\mathbf{W}_{j,1} = \mathbf{W}_j$. We also know that the spaces $\mathbf{W}_{j-1,2n}$ and $\mathbf{W}_{j-1,2n+1}$ are orthogonal and their direct sum is $\mathbf{W}_{j,n}$. Therefore the leaves of every connected binary subtree[9] of the wavelet packet tree correspond to an orthogonal basis of the initial space.

Each of the bases encoded in the wavelet packet corresponds to a dyadic tiling of the phase plane in Heisenberg boxes of area one. They provide a much richer time/frequency analysis.

Each of the spaces is generated by the integer shifts of a wavelet function at scale j and frequency n. More precisely, let $\omega_{j,k,n}(t) = 2^{-j/2}\omega_n(2^{-j}t - k)$, where $n \in \mathbb{N}$, $j, k \in \mathbb{Z}$, and

$$
\begin{aligned}
\omega_{2n}(t) = \sqrt{2}\sum h_k \omega_n(2t - k), \quad \omega_0 = \phi, \\
\omega_{2n+1}(t) = \sqrt{2}\sum g_k \omega_n(2t - k), \quad \omega_1 = \psi,
\end{aligned}
$$

where $\{h_k\}$ and $\{g_k\}$ are the low and high-pass filters of the MRA.

Then $\mathbf{W}_{j,n} = \mathrm{span}\{\omega_{j,k,n} : k \in \mathbb{Z}\} = \{f = \sum_k a_{j,k,n} w_{j,k,n} : \sum_k |a_{j,k,n}|^2 < \infty\}$.

Example 18 *For the Haar function, the corresponding wavelet packet, is described by the equations,*

$$
\begin{aligned}
\omega_{2n}(t) &= \omega_n(2t) + \omega_n(2t - 1) \\
\omega_{2n+1}(t) &= \omega_n(2t) - \omega_n(2t - 1).
\end{aligned}
$$

The functions so obtained are the Walsh functions, *which are step functions that play the role of the sines and cosines.*

[9] Starting from the "top" node ($\mathbf{W}_{0,0}$), we either allow the node to have two off-springs or none. Then we repeat with the off-springs if there are any. The nodes that have no off-springs are the *leaves*.

Exercise 47 *Identify all possible bases for the Haar packet binary tree for 3 levels. How many bases can you find? Draw corresponding phase plane diagrams. In particular, draw the Walsh functions.*

Exercise 48 *Who are the Walsh functions in the finite dimensional case?*

Thus a signal of length $N = 2^J$ can be decomposed in 2^N different ways, the number of binary subtrees of a complete binary tree of depth J. This is a large number, and one would like to search efficiently in the tree to obtain the best basis with respect to some criteria.

Functionals verifying an additive-type property are well suited for this type of searches. Coifman and Wickerhauser introduced a number of such functionals [CW], among them some *entropy criteria*. Given a signal s and $(s_i)_i$ its coefficients in an orthonormal basis. The entropy E must be an additive cost function such that $E(0) = 0$ and $E(s) = \sum_i E(s_i)$. There are fast algorithms that allow to search in the wavelet packet tree for the orthonormal basis that minimizes some given entropies (there are four of them encoded in Matlab). Furthermore the search can be performed in $O(N \log(N))$ operations. This is all implemented in the Wavelet toolbox.

The wavelet packet and *cosine packet* libraries create large libraries of orthogonal bases, all of which have fast algorithms. The Fourier and Wavelet basis are just examples in this time/frequency library, so are Gabor like basis.

4.6 Wavelets in 2-D

There is a standard procedure to construct bases in 2-D space from given bases in 1-D, the *tensor product*. In particular, given a wavelet basis $\{\psi_{j,k}\}$ in $L^2(\mathbb{R})$, the family of tensor products

$$\psi_{j,k;i,n}(x,y) = \psi_{j,k}(x)\psi_{i,n}(y), \quad j,k,i,n \in \mathbb{Z},$$

is an orthonormal basis in $L^2(\mathbb{R}^2)$. Unfortunately we have lost the multiresolution structure. Notice that we are mixing up scales in the above process, that is the scaling parameters i, j can be anything.

Exercise 49 *What would be the trigonometric basis in $L^2([0,1]^2)$?*
How about the finite dimensional trigonometric basis in 2-D?

We would like to use this idea but at the level of the approximation spaces $\mathbf{V}_j$ in the MRA. For each scale j, the family $\{\phi_{j,k}\}_k$ is an orthonormal basis of $\mathbf{V}_j$. Consider the tensor products of these functions, $\phi_{j,k,n}(x,y) = \phi_{j,k}(x)\phi_{j,n}(y)$, then let $\mathcal{V}_j$ be the closure in $L^2(\mathbb{R}^2)$ of the linear span of those functions, that is,

$$\mathcal{V}_j = \mathbf{V}_j \otimes \mathbf{V}_j = \{f(x,y) = \sum_{n,k} a_{j,n,k}\phi_{j,k,n}(x,y) : \sum_{n,k} |a_{j,n,k}|^2 < \infty\}.$$

Notice that we are not mixing scales at the level of the MRA. It is not hard to see that the spaces $\mathcal{V}_j$ form an MRA in $L^2(\mathbb{R}^2)$ with scaling function

$$\phi(x,y) = \phi(x)\phi(y).$$

This means that the integer shifts $\{\phi(x-k, y-n) = \phi_{0,k,n}\}_{k,n\in\mathbb{Z}}$ form an orthonormal basis of $\mathcal{V}_0$, consecutive approximation spaces are connected via scaling by 2 on both variables, and the other conditions are clear.

The orthogonal complement of $\mathcal{V}_j$ in $\mathcal{V}_{j+1}$ is the space $\mathcal{W}_j$ which can be seen is the direct sum of three tensor products, namely,

$$\mathcal{W}_j = (\mathbf{W}_j \otimes \mathbf{W}_j) \oplus (\mathbf{W}_j \otimes \mathbf{V}_j) \oplus (\mathbf{V}_j \otimes \mathbf{W}_j).$$

Therefore *three wavelets* are necessary to span the detail spaces,

$$\psi^d(x,y) = \psi(x)\psi(y), \quad \psi^v(x,y) = \psi(x)\phi(y), \quad \psi^h(x,y) = \phi(x)\psi(y),$$

where d stands for diagonal, v for vertical, and h for horizontal (the reason for such names is that each of the subspaces will somehow favor details in those directions).

In higher dimensions the same works. There will be one scaling function, and $2^n - 1$ wavelets, where n is the dimension.

Exercise 50 *Describe a* 3*-dimensional MRA. These are useful for video compression.*

This construction has the advantage that the bases are separable, implementing the fast two dimensional wavelet transform is not difficult. In fact it can be done by successively applying the one dimensional FWT. The disadvantage is that the analysis is very axis-dependent, which might not be desirable for certain applications.

Example 19 (2-D Haar basis) *The scaling function is the characteristic function of the unit cube,*

$$\phi(x,y) = \chi_{[0,1]^2}(x,y).$$

The following pictures should suffice to understand the nature of the 2-D Haar wavelets and scaling function:

$$\begin{array}{|c|c|}\hline 1 & 1 \\ \hline 1 & 1 \\ \hline \end{array} \qquad \begin{array}{|c|c|}\hline -1 & 1 \\ \hline 1 & -1 \\ \hline \end{array} \qquad \begin{array}{|c|c|}\hline 1 & -1 \\ \hline 1 & -1 \\ \hline \end{array} \qquad \begin{array}{|c|c|}\hline -1 & -1 \\ \hline 1 & 1 \\ \hline \end{array}$$

$$\phi(x,y) \qquad\quad \psi^d(x,y) \qquad\quad \psi^h(x,y) \qquad\quad \psi^v(x,y)$$

Exercise 51 *Describe the* 2*-dimensional discrete Haar MRA.*

There are *non-separable* two dimensional MRA's. The most famous one corresponds to an analogue of the Haar basis. The scaling function is the characteristic function of a two dimensional set. It turns out that the set has to be rather complicated, in fact it is a self-similar set with fractal boundary, the so-called *twin dragon*.

4.7 Basics of compression and denoising

One of the main interests in signal and image processing is to be able to code the information with as few data as possible. This allows for rapid transmission, etc. In the presence of noise, one wants to separate the noise from the signal, and one would like to have a basis that concentrates the signal in a few large coefficients and delegates the noise to very small coefficients.

In both the deterministic and the noisy case, the steps to follow are:

- Transform the data, find coefficients in a given basis.
- Threshold the coefficients (essentially one keeps the large ones and discards the small ones, information is lost in this step, so perfect reconstruction is not possible).
- Reconstruct with the thresholded coefficients, and hope that you have obtained a good approximation to your original signal (compressed signal), or you have successfully denoised it.

Wavelet basis are good for decorrelating coefficients. They are also good for denoising in the presence of *white noise*.

The crudest approach would be to use the projection into an approximation space as your compressed signal, discarding all the details after certain scale j,

$$P_j f = \sum_k \langle f, \phi_{j,k}\rangle \phi_{j,k}.$$

The noise is usually concentrated in the finer scales (higher frequencies!), so this approach would denoise but at the same time it will remove many of the sharp features of the signal that were encoded in the finer wavelet coefficients. A more refined technique is required, called *thresholding*. There are different thresholding techniques. The most popular are *hard thresholding* (it is a keep or toss thresholding), and *soft thresholding* (the coefficients are attenuated following a linear scheme). There is also the issue about thresholding individual coefficients or *block-thresholding*. How to select the threshold is another issue. In the denoising case, there are some thresholding selection rules which are justified by probability theory (basically the law of the large numbers), and are used widely by statisticians. Both thresholding and threshold selection principles are encoded in Matlab.

In traditional approximation theory there are two possible methods, *linear* and *non-linear* approximation.

- *Linear approximation:* This refers to selecting a priori N elements in the basis and projecting onto the subspace generated by those elements, regardless of the function that is being approximated. This is a linear scheme,

$$P_N^l f = \sum_{n=1}^{N} \langle f, \psi_n\rangle \psi_n.$$

- *Non-linear approximation:* This approach chooses the basis elements depending on the function, for example the N basis elements are chosen so that the coefficients are the largest in size for the particular function. This time the chosen basis elements will depend on the particular function to be approximated,

$$P_N^{nl} f = \sum_{n=1}^{N} \langle f, \psi_{n,f} \rangle \psi_{n,f}.$$

The non-linear approach has proven quite successful. There is a lot more information about these issues in [Mall98, Ch. 9-10].

5 The Hilbert Transform

The Hilbert transform is the prototypical example of a singular integral operator. It is given formally by the principal value integral:

$$Hf(x) = \text{p.v.}\frac{1}{\pi}\int \frac{f(y)}{x-y}dy := \lim_{\epsilon\to 0}\frac{1}{\pi}\int_{|x-y|>\epsilon}\frac{f(y)}{x-y}dy.$$

One can compute the Fourier transform of the Hilbert transform, at least when applied to very smooth and compactly supported functions, and one obtains:

$$(Hf)^{\wedge}(\xi) = \int_{\mathbb{R}} Hf(x)e^{-2\pi i x\xi}dx = -i\,\mathrm{sgn}(\xi)\hat{f}(\xi). \tag{35}$$

Here we define $\mathrm{sgn}(\xi) = -1$ if $\xi < 0$, $\mathrm{sgn}(\xi) = 1$ if $\xi > 0$, and $\mathrm{sgn}(0) = 0$. See Exercise 53 for an idea on how to prove this formula.

This automatically shows that the Hilbert transform is an isometry

$$\|Hf\|_2 = \|(Hf)^{\wedge}\|_2 = \|\hat{f}\| = \|f\|_2,$$

where Plancherel's identity has been used twice, at least on a dense subset of $L^2(\mathbb{R})$. Notice also that from (35), we conclude that $H^2 = -I$ (minus the identity operator), since

$$(H^2 f)^{\wedge} = -i\,\mathrm{sgn}(\xi)\hat{H}f = (-i\,\mathrm{sgn}(\xi))^2\hat{f} = -\hat{f}.$$

Exercise 52 *Show that if $Ff = \hat{f}$, then F^4 is the identity operator, at least in the Schwartz class.*

It is desirable to obtain alternative proofs of the fact that the Hilbert transform is an isometry, independent of the Fourier transform. Time permitting we will give an alternative proof (or at least the main ideas) of a seemingly less powerful statement, namely, there exists a positive constant C, such that for all $f \in L^2(\mathbb{R})$,

$$\|Tf\|_2 \le C\|f\|_2.$$

This inequality means that the operator T is *bounded* in L^2. Similarly we can define boundedness in L^p: an *operator T is bounded in $L^p(\mathbb{R})$*[10], $1 \le p \le \infty$, if there exists a positive constant C, such that for all $f \in L^p(\mathbb{R})$,

$$\|Tf\|_p \le C\|f\|_p.$$

[10] For linear operators, boundedness is equivalent to continuity. In this case, T is bounded in $L^p(\mathbb{R})$ if and only if

$$\|Tf_n - Tf\|_p \to 0, \quad \text{whenever} \quad \|f_n - f\|_p \to 0.$$

Notice that the Hilbert transform is given by convolution with the distributional kernel $k(x) = \text{p.v.}\frac{1}{x}$. Had the kernel been an integrable function, then the *integral operator* T, given by convolution against the kernel,

$$Tf(x) = k * f(x) = \int k(x-y)f(y)dy$$

would had been automatically bounded in L^p for all $1 \leq p \leq \infty$, by *Young's inequality*,

$$\|f * k\|_p \leq \|k\|_1 \|f\|_p,$$

which we will verify at the end of the lecture. Unfortunately the Hilbert transform kernel is not integrable, but still the Hilbert transform is bounded in $L^p(\mathbb{R})$ for all $1 < p < \infty$; and although it is not bounded at the endpoints $p = 1$ and $p = \infty$; there are appropriate substitutes.

5.1 Some History

Why did mathematicians get interested in the Hilbert transform? Here are a couple of classical problems where the Hilbert transform appeared naturally.

5.1.1 Connection to complex analysis

Consider a real valued function $f \in L^2(\mathbb{R})$ and let $F(z)$ be twice its analytic extension to the upper half plane $\mathbb{R}^2_+ = \{z = x + iy : y > 0\}$, suitably normalized. $F(z)$ can be explicitly computed by means of the well known *Cauchy integral formula*:

$$F(z) = \frac{1}{\pi i} \int_{\mathbb{R}} \frac{f(t)}{z-t} dt, \qquad z \in \mathbb{R}^2_+.$$

Notice the resemblance with the Hilbert transform. No principal value is needed here since the singularity is never achieved. By separating the real and imaginary parts of the kernel, one can obtain explicit formulae for the real and imaginary parts of $F(z) = u(z) + iv(z)$ in terms of convolutions with the *Poisson* and *Conjugate Poisson kernels*: $u(x+iy) = f * P_y(x)$, $v(x+iy) = f * Q_y(x)$, that we encounter in the first lecture. The function u is the *harmonic extension* of f to the upper-half plane, and the function v is its *harmonic conjugate*.

Exercise 53 *Show that the Poisson kernel is given by* $P_y(x) = \frac{1}{\pi}\frac{y}{x^2+y^2}$*, and the conjugate Poisson kernel by* $Q_y(x) = \frac{1}{\pi}\frac{x}{x^2+y^2}$*. You already showed in exercise* 15 *that for each* $y > 0$, $\hat{Q}_y(\xi) = -i\,\text{sgn}(\xi)e^{-2\pi|y\xi|}$*, therefore as* $y \to 0$, $\hat{Q}_y(\xi)$ *approaches* $-i\text{sgn}(\xi)$*. On the other hand, as* $y \to 0$, Q_y *approaches the principal value distribution p.v.*$\frac{1}{\pi x}$*, which means in this case that for all test function* $\phi \in \mathcal{S}$,

$$\lim_{y\to 0} \int Q_y(x)\phi(x)dx = \lim_{\epsilon\to 0} \int_{|x|>\epsilon} \frac{\phi(x)}{x} dx.$$

The Poisson kernel is an example of an approximation of the identity that we already discussed in the first lecture. As such, the limit as $y \to 0$ of $u = P_y * f$ is f in the L^2-sense and almost everywhere. On the other hand, as $y \to 0$, $v = Q_y * f$ approaches the Hilbert transform Hf in $L^2(\mathbb{R})$.

5.1.2 Connection to Fourier series

For functions integrable on $\mathbb{T} = [0, 1]$, the n-th Fourier coefficient is well defined by the formula

$$\hat{f}(n) = \int_0^1 f(x) e^{-2\pi i n x} dx.$$

Since $L^2(\mathbb{T}) \subset L^1(\mathbb{T})$, this is also well defined for square integrable functions. We already know that the trigonometric system $\{e^{2\pi i n x}\}_{n \in \mathbb{Z}}$ is an orthonormal complete system in $L^2(\mathbb{T})$; therefore the following reconstruction and isometry formulae hold in L^2,

$$f(x) = \sum_{n \in \mathbb{Z}} \hat{f}(n) e^{2\pi i n x}, \qquad \|f\|^2_{L^2(\mathbb{T})} = \sum_{n \in \mathbb{Z}} |\hat{f}(n)|^2.$$

The N-th partial sum is given by

$$S_N f(x) = \sum_{|n| \leq N} \hat{f}(n) e^{2\pi i n x}.$$

In the XIX century, mathematicians asked for which 2π-periodic functions f would it be true that

$$\lim_{N \to \infty} S_N f(x) = f(x)$$

at a given point $x \in \mathbb{T}$? Some partial answers were given, and more than continuity at the point was always required (eg. Dini's condition). In 1889, Du Bois Raymond found a continuous function whose partial Fourier sum diverges at a point. Summation methods were introduced to counter these "pathologies". You discussed many of these issues the first week. With the advent of measure theory and L^p spaces, new questions were formulated.

- **Is there convergence a.e.?** That is, does there exist an exceptional set E of measure zero, such that for all x not included in E

$$\lim_{N \to \infty} S_N f(x) = f(x) \ ?$$

- **Is there convergence in the L^p sense?** That is, does

$$\lim_{N \to \infty} \|S_N f(x) - f(x)\|_p = 0 \ ?$$

The second question is answered positively for $1 < p < \infty$ and it is a consequence of the boundedness of the Hilbert transform in such L^p's. The first question is much more difficult, for $p = 2$ the positive answer was given by L. Carleson in a celebrated paper published in 1965, see [Car] (settling the question for periodic continuous functions which had remained open until then); two years later, R. Hunt extended the result for the remaining p's, $1 < p < \infty$, see [Hu]. The case $p = 1$ had been ruled out by Kolmogorov's famous example of an integrable function whose Fourier series diverges everywhere, see [Kol].

By a limiting procedure on the unit disc, similar to the one described in the upper half plane, one can conclude that the boundary value of the harmonic conjugate of the harmonic extension of a periodic, real-valued, continuously differentiable function f on $\mathbb{T}$ is given by

$$\tilde{H}f(x) = \text{p.v.}\frac{1}{\pi}\int_0^1 f(t)\cot\big(\pi(x-t)\big)\,dt. \tag{36}$$

Here we are identifying the function $f(x)$ with $f(z)$, for $z = e^{2\pi i x}$.

The singularity at the diagonal is comparable to that of the Hilbert transform. This would be the appropriate analogue of the Hilbert transform on the unit circle. The analogy gets reinforced when we notice that on Fourier side,

$$(\tilde{H}f)^\wedge(n) = -i\,\text{sgn}(n)\hat{f}(n).$$

We could have defined $\tilde{H}$ by this formula, as many authors do, and then worked backwards to compute the inverse Fourier transform to obtain (36). This is an example of a *Fourier multiplier* on the circle. These operators are bounded linear transformations on $L^p(\mathbb{T})$ given by the formula

$$(Tf)^\wedge(n) = m(n)\hat{f}(n),$$

where the sequence $\{m(n)\}_{n\in\mathbb{Z}}$ is the *multiplier.*

Both H and $\tilde{H}$ share similar boundedness properties. In particular, $\tilde{H}$ is bounded in $L^p(\mathbb{T})$ for $1 < p < \infty$, that is,

$$\|\tilde{H}f\|_{L^p(\mathbb{T})} \leq C\|f\|_{L^p(\mathbb{T})}.$$

Note that the partial Fourier sums of a nice function are also Fourier multipliers,

$$(S_N f)^\wedge(n) = \chi_{|k|\leq N}(n)\hat{f}(n).$$

Exercise 54 *Check that*

$$\chi_{|k|\leq N}(n) = \frac{1}{2}(\text{sgn}(n-N) - \text{sgn}(n+N)).$$

Furthermore remember that the Fourier transform maps modulations into translations. Denote $M_N f(\theta) = f(x)e^{2\pi i\theta N}$, and check that

$$i(M_N\tilde{H}M_{-N})^\wedge(n) = \text{sgn}(n-N)\hat{f}(n), \quad i(M_{-N}\tilde{H}M_N)^\wedge(n) = \text{sgn}(n+N)\hat{f}(n).$$

The exercise implies, after applying the inverse Fourier transform, that

$$S_N = \frac{i}{2}(M_N \tilde{H} M_{-N} - M_{-N} \tilde{H} M_N).$$

¿From the boundedness in $L^p(\mathbb{T})$ of $\tilde{H}$, we conclude that the S_Ns are uniformly (in N) bounded in L^p, for each $1 < p < \infty$. We also know that the trigonometric polynomials are dense in $L^p(\mathbb{T})$. Given one of them, $f(x) = \sum_{|m| \leq M} a_m e^{2\pi i x m}$, then for $N \geq M$, $S_N f = f$. The partial sum operators S_N converge to the identity operator on a dense subset of $L^p(\mathbb{T})$. We can now deduce from the Uniform Boundedness Principle Stated in Section 3.3.4, that

$$\lim_{N \to \infty} \|S_N f - f\|_p = 0.$$

Therefore the convergence in $L^p(\mathbb{T})$ of the partial Fourier sums is a consequence of the boundedness of the Hilbert transform in those spaces.

Exercise 55 *Given $f(x) = \sin(2\pi x)$ on $[0, 1]$, find $\tilde{H} f$.*

Fourier multipliers correspond to operators given by convolution with a distributional kernel. Their generalizations, *singular integral operators* and *Fourier integral operators*, have many applications, and are still the object of intense study today.

5.2 Weak $(1, 1)$

When $p = 1$, one can get away with a weaker notion of boundedness, which requires notions of measure theory. Suppose, $g \in L^1(\mathbb{R})$, and let

$$E_\lambda(g) = \{x \in \mathbb{R} : |g(x)| > \lambda\} \subset \mathbb{R}.$$

Assume T is a linear bounded operator in $L^1(\mathbb{R})$, namely,

$$\|Tf\|_1 = \int_{\mathbb{R}} |Tf(x)| dx \leq C \|f\|_1.$$

Even if we are not fully acquainted with the Lebesgue integral, our experience with Riemann integration tells us that the following inequality is plausible,

$$\int_{\mathbb{R}} |Tf(x)| dx \geq \int_{E_\lambda(Tf)} |Tf(x)| dx.$$

Now observe that by definition, if $x \in E_\lambda(Tf)$, then $|Tf(x)| > \lambda$, hence we can also believe[11] that

$$\int_{E_\lambda(Tf)} |Tf(x)| dx \geq \int_{E_\lambda(Tf)} \lambda \, dx = \lambda \int_{E_\lambda(Tf)} 1 \, dx = \lambda \int_{\mathbb{R}} \chi_{E_\lambda(Tf)}(x) \, dx.$$

[11] The properties of the integral that we are using are: (i) for $f \geq 0$, and $A \subset B$ then $\int_A f \leq \int_B f$; (ii) if $f \leq g$ on A, then $\int_A f \leq \int_B g$; (iii) if $\lambda \in \mathbb{R}$, then $\int_A \lambda f = \lambda \int_A f$; (iv) $\int_A 1 dx = \int \chi_A(x) dx$.

Finally if E_λ were an interval, then this last integral is equal to its length. It turns out that if $E_\lambda(Tf)$ is a *measurable set*, one can extend the notion of "length" to such sets, the so-called *Lebesgue measure*. The Lebesgue measure of a measurable set A is denoted by $m(A)$, and it coincides with the Lebesgue integral[12] of χ_A on $\mathbb{R}$, i.e.

$$m(A) = \int_A 1\,dx = \int_{\mathbb{R}} \chi_A(x)\,dx.$$

What we have attempted to show is that if T is bounded $L^1(\mathbb{R})$, then there is a constant $C \geq 1$ such that,

$$m(\{x \in \mathbb{R} : |Tf(x)| > \lambda\}) \leq \frac{C}{\lambda}\|f\|_1, \tag{37}$$

An operator that satisfies (37) is said to be of *weak type* $(1,1)$. We have just shown that bounded in $L^1(\mathbb{R})$ implies weak $(1,1)$; but the converse is, in general, false.

Exercise 56 *Show* Chebyshev's inequality*, that if* $f \in L^p(\mathbb{R})$*, then*

$$m(\{x \in \mathbb{R} : |f(x)| > \lambda\}) \leq \left(\frac{\|f\|_p}{\lambda}\right)^p.$$

An operator is said to be of *weak-type* (p,p), if there is a $C > 0$ such that for all $\lambda > 0$,

$$m(\{x \in \mathbb{R} : |Tf(x)| > \lambda\}) \leq \left(\frac{C\|f\|_p}{\lambda}\right)^p. \tag{38}$$

Exercise 57 *Show that if T is bounded in L^p then it is of weak-type (p,p).*

Theorem 8 *The Hilbert transform is of weak type* $(1,1)$*, but not bounded in* $L^1(\mathbb{R})$*.*

Exercise 58 shows that H cannot be bounded in $L^1(\mathbb{R})$, nor in $L^\infty(\mathbb{R})$. We will not show here that H is of weak-type (1,1). This is a consequence of the celebrated *Calderón-Zygmund decomposition* which, if we had time deserves to be taught in a course like this one. It is an example of another technique widely used in modern harmonic analysis, *stopping time techniques*. Bounded functions are mapped into a larger space *Bounded Mean Oscillation*, BMO, which contains unbounded functions like $\ln|x|$. We will not show these facts here.

This behavior is shared by a large class of very important operators the so-called *Calderón-Zygmund Singular Integral Operators*. The departure point of the Calderón-Zygmund theory is an a priori L^2-estimate; everything else unfolds from there. Having means other than Fourier analysis to obtain such L^2-estimate becomes important in other contexts. We will address this issue for the Hilbert transform in the last lecture.

[12] We are not defining this in these notes. The difference between Riemann and Lebesgue integrals is like adding loose change one coin at a time (Riemann) vs collecting them by denomination (pennies, dimes, quarters, etc), multiplying each denomination by the number of coins of that denomination, and then adding up (Lebesgue).

Exercise 58 *Find the Hilbert transform of* $\chi_{[0,1]}$, *observe that it is not an integrable function nor a bounded function, despite the fact that the characteristic function is both. Try to justify the following inequality,*

$$m\{x \in \mathbb{R} : |H\chi_{[0,1]}(x)| \geq \lambda\} \leq \frac{C}{\lambda}.$$

We mentioned the Lebesgue Differentiation Theorem (23) in Section 3.3.4. We know the theorem holds for continuous functions as a consequence of the Fundamental Theorem of Calculus. A related operator, which is defined for all $x \in \mathbb{R}$ (possibly to be ∞), is the *Hardy-Littlewood Maximal operator*,

$$Mf(x) = \sup_{h>0} \frac{1}{2h} \int_{x-h}^{x+h} |f(t)|\, dt.$$

The maximal function is also of weak type $(1,1)$, this is also a consequence of the Calderón-Zygmund decomposition Lemma. This fact, together with the knowledge of the existence of the limit on a dense subset, allows us to show the Lebesgue Differentiation Theorem for integrable functions. Maximal functions appear all over in harmonic analysis as controllers for other operators, and understanding their boundedness properties is very important.

5.2.1 The distribution function and L^p

Suppose f is a function such that $E_\lambda(f)$ is measurable for all $\lambda \geq 0$ (These are exactly the *measurable functions*. Step functions are always measurable). Then we can define the *distribution function*, d_f, of f, on $[0,\infty)$ by

$$d_f(\lambda) = m\big(E_\lambda(f)\big) = m\{x \in \mathbb{R} : |f(x)| > \lambda\}.$$

Notice that d_f is always a decreasing function.

Example 20 *Let* $f = 2\chi_{[-1,2]} - \chi_{[3,5]} + 7\chi_{[10,11]}$. *Then*

$$d_f = 6\chi_{[0,1)} + 4\chi_{[1,2)} + \chi_{[2,7)}.$$

Exercise 59 *Let* $\{I_k : k = 1, \ldots n\}$ *be disjoint intervals. Let* f *be the step function given by* $f = \sum_{k=1}^{n} a_k \chi_{I_k}$. *Find its distribution function* d_f.

Exercise 60 *Show that* $d_f(\lambda) = d_{\tau_h f}(\lambda)$ *for all* $h \in \mathbb{R}$, *and for* f *a step functions like the ones in exercise* 59. *That is the distribution function of* f *coincides with the distribution function of any of its translates. (This holds for all measurable functions.)*

The previous exercise shows that the distribution function provides information about the size of f, but not about its behavior near any given point. However, it does provide sufficient information to compute L^p-norms, which are invariant under translation.

Theorem 9 *For $f \in L^p(\mathbb{R})$, $1 \le p < \infty$ we have that*

$$\|f\|_p^p = p \int_0^\infty \lambda^{p-1} d_f(\lambda) d\lambda.$$

Proof: By definition of d_f, we have,

$$p \int_0^\infty \lambda^{p-1} d_f(\lambda) d\lambda = p \int_0^\infty \int_{\{x:|f(x)|>\lambda\}} \lambda^{p-1} dx\, d\lambda.$$

Interchanging the order of integration, which is allowed by Fubini, we get

$$\int_{\mathbb{R}} \int_0^{|f(x)|} p\lambda^{p-1} d\lambda\, dx = \int_{\mathbb{R}} |f(x)|^p dx = \|f\|_p^p.$$

$\diamond$

Exercise 61 *Make sure you believe this proof at least for the case of step functions like the ones in Exercise* 59.

Exercise 62 *Show that for any increasing continuously differentiable function ϕ on $[0,\infty)$ we have*

$$\int_{\mathbb{R}} \phi(|f|) dx = \int_0^\infty \phi'(\lambda) d_f(\lambda) d\lambda.$$

5.3 Interpolation

We are going to use this section to state some very useful theorems that go under the generic name of *interpolation theorems*. There are books devoted to the subject. We will present a proof in a very simplified scenario. Basically these theorems tell us that if we know the boundedness properties of a linear operator at two end-points, then we automatically obtain boundedness at many other points.

5.3.1 The Marcinkiewicz Interpolation Theorem

Theorem 10 *Let T be a linear operator defined on $L^p(\mathbb{R}) + L^q(\mathbb{R})$, where $1 \le p < q \le \infty$. Assume T is of weak-type (p,p) and (q,q), that is,*

$$\sup_{\lambda>0} \lambda \big(d_{Tf}(\lambda)\big)^{1/p} \le A_p \|f\|_p, \quad \textit{and} \quad \sup_{\lambda>0} \lambda \big(d_{Tf}(\lambda)\big)^{1/q} \le A_q \|f\|_q.$$

Then T is bounded in $L^r(\mathbb{R})$, for all $p < r < q$, moreover,

$$\|Tf\|_r \le 2 \left(\frac{r}{r-p} + \frac{r}{q-r} \right)^{1/r} A_p^t A_q^{1-t} \|f\|_r,$$

where $t \in (0,1)$ satisfies $\frac{1}{r} = \frac{t}{p} + \frac{1-t}{q}$.

We will only prove the theorem in the case T is the identity operator, which is an statement on the nature of $L^p(\mathbb{R})$ spaces. The proof of the Marcinkiewicz interpolation theorem can be found in most books in harmonic analysis or real analysis. If we replace the weak-type hypothesis by boundedness on $L^p(\mathbb{R})$ and $L^q(\mathbb{R})$, one can get the same result with constant one, that is, T is bounded in $L^r(\mathbb{R})$, and

$$\|Tf\|_r \leq A_p^t A_q^{1-t}\|f\|_r.$$

This can be seen as a corollary of the Riesz-Thorin interpolation theorem in the next section, which we will state but not prove either.

Proposition 1 *If $f \in L^p(\mathbb{R})\bigcap L^q(\mathbb{R})$, $1 \leq p < q \leq \infty$, then $f \in L^r(\mathbb{R})$ for all $p < r < q$, moreover,*

$$\|f\|_r \leq \left(\frac{r}{r-p} + \frac{r}{q-r}\right)^{1/r} \|f\|_p^t \|f\|_q^{1-t},$$

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

Proof of Proposition 1: Assume $q < \infty$. By Chebyshev's inequality (see exercise 56), if $f \in L^p(\mathbb{R})\bigcap L^q(\mathbb{R})$, then,

$$d_f(\lambda) \leq \min\left\{\frac{\|f\|_p^p}{\lambda^p}, \frac{\|f\|_q^q}{\lambda^q}\right\}.$$

Let $B = \left(\frac{\|f\|_q^q}{\|f\|_p^p}\right)^{1/(q-p)}$, and note that if $0 \leq \lambda \leq B$, then the above minimum is $\|f\|_p^p\lambda^{-p}$, and if $B \leq \lambda \leq \infty$, then the minimum is $\|f\|_q^q\lambda^{-q}$. We can now estimate the L^r-norm of f. By Theorem 9, we have that

$$\begin{aligned}
\|f\|_r^r &= r\int_0^\infty \lambda^{r-1} d_f(\lambda)d\lambda \\
&\leq r\int_0^\infty \lambda^{r-1}\min\left\{\frac{\|f\|_p^p}{\lambda^p}, \frac{\|f\|_q^q}{\lambda^q}\right\} d\lambda \\
&= r\int_0^B \lambda^{r-1-p}\|f\|_p^p d\lambda + r\int_B^\infty \lambda^{r-1-q}\|f\|_q^q d\lambda \\
&= \frac{r}{r-p}\|f\|_p^p B^{r-p} + \frac{r}{q-r}\|f\|_q^q B^{r-q} \\
&= \left(\frac{r}{r-p} + \frac{r}{q-r}\right)\|f\|_p^{p\frac{(q-r)}{q-p}}\|f\|_q^{q\frac{(r-p)}{q-p}}.
\end{aligned}$$

Finally just note that $t = \frac{p}{r}\frac{(q-r)}{q-p}$.

The case when $q = \infty$ is left as an exercise to the reader. $\diamond$

The inequality in the proposition actually holds with constant one, see Exercise 64. We have presented this proof because it is similar in spirit to the proof of the interpolation theorem, where the hypotheses involve weak-type inequalities. Notice that all we needed in this proof was Chebyshev's inequality, which always holds for functions in $L^p(\mathbb{R})$. But we could have asked the function to satisfy "weak" L^p and L^q-estimates, that is, replace $\|f\|_p$ by the quantity

$$[f]_p = \sup_{\lambda>0} \lambda (d_f(\lambda))^{1/p},$$

and $\|f\|_q$ by $[f]_q$, and we still would have concluded that if both $[f]_p$, and $[f]_q$ are finite, then $f \in L^r$ and

$$\|f\|_r \leq \left(\frac{r}{r-p} + \frac{r}{q-r} \right)^{1/r} [f]_p^t [f]_q^{1-t}, \quad \text{where} \quad \frac{1}{r} = \frac{t}{p} + \frac{1-t}{q}.$$

5.3.2 L^p boundedness of the Hilbert transform

Corollary 1 *H is bounded in $L^p(\mathbb{R})$ for $1 < p < \infty$.*

Proof: We mentioned (although we did not prove it!) that H is of weak-type (1,1), and is also bounded in $L^2(\mathbb{R})$, then the Marcinkiewicz interpolation theorem implies that H is bounded in $L^p(\mathbb{R})$ for all $1 < p < 2$. The adjoint of H is $-H$, and it turns out that if H is bounded in $L^p(\mathbb{R})$, then its adjoint is bounded in $L^{p'}(\mathbb{R})$, where p' is the dual exponent of p, that is, $\frac{1}{p} + \frac{1}{p'} = 1$. This duality argument now takes care of the range $2 < p < \infty$. $\diamond$

5.3.3 The Riesz-Thorin Interpolation Theorem

Here is a variant of the interpolation theorem above, where the operator does not necessarily map a given space into itself. It is not exactly a generalization of the Marcinkiewicz interpolation theorem because it does not deal with operators of weak-type, but it allows for different L^p spaces as inputs and outputs of the operator. The constant that appears is simpler than in the Marcinkiewicz theorem, is just one. This theorem has as almost immediate corollaries some of the most famous inequalities in analysis.

Theorem 11 *Let $1 \leq p_1, p_2, q_1, q_2 \leq \infty$, and for $0 \leq t \leq 1$ define r_1 and r_2 by*

$$\frac{1}{r_1} = \frac{t}{p_1} + \frac{1-t}{q_1}, \qquad \frac{1}{r_2} = \frac{t}{p_2} + \frac{1-t}{q_2}.$$

It T is a linear operator from $L^{p_1} + L^{p_2}$ into $L^{q_1} + L^{q_2}$, such that

$$\|Tf\|_{p_2} \leq A_1 \|f\|_{p_1} \quad \textit{for all} \;\; f \in L^{p_1},$$

and

$$\|Tf\|_{q_2} \leq A_2 \|f\|_{q_1} \quad \textit{for all} \;\; f \in L^{q_1},$$

then

$$\|Tf\|_{r_2} \leq A_1^t A_2^{1-t} \|f\|_{r_1} \quad \textit{for all} \;\; f \in L^{r_1}.$$

The proof of this theorem uses the so-called "three-lines" theorem for analytic functions, it can be found in any harmonic analysis textbook (e.g. [StWe, Ch. 5], [Gr, Ch. 1]).

5.3.4 An Inequalities' Festival

Corollary 2 (Hausdorff-Young Inequality) *If $f \in L^p(\mathbb{R})$, $1 \leq p \leq 2$, then $\hat{f} \in L^q$, for $\frac{1}{p} + \frac{1}{q} = 1$, and*

$$\|\hat{f}\|_q \leq \|f\|_p.$$

Exercise 63 *Prove the Hausdorff-Young inequality by interpolating between Plancherel, and the $L^1 - L^\infty$ inequality, $\|\hat{f}\|_\infty \leq \|f\|_1$, discussed briefly in the first lecture. Notice that it can be deduced from the simple estimate*

$$|\hat{f}(x)| \leq \int_{\mathbb{R}} |f(t)|dt = \|f\|_1.$$

Corollary 3 (General Hölder's inequality) *Given $f \in L^p(\mathbb{R})$, and $g \in L^q(\mathbb{R})$, then $fg \in L^r(\mathbb{R})$, where $\frac{1}{r} = \frac{1}{p} + \frac{1}{q}$, and*

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

Proof: Fix $1 \leq q \leq \infty$, and $g \in L^q$. Show first that if $p = \infty$, then $r = q$ and the following inequality holds,

$$\|fg\|_q \leq \|f\|_\infty \|g\|_q.$$

Next remember *Hölder's inequality*, which corresponds to $r = 1$, $p = q'$, where $\frac{1}{q} + \frac{1}{q'} = 1$,

$$\|fg\|_1 = \int_{\mathbb{R}} |f(x)g(x)|dx \leq \|f\|_{q'} \|g\|_q. \tag{39}$$

Now apply the Riesz-Thorin interpolation theorem with $p_1 = q'$, $p_2 = \infty$, $r_1 = 1$, $r_2 = q$, and the linear operator is given by multiplication by g, $Tf = fg$. $\diamond$

Exercise 64 *Show that Proposition 1 holds with constant one. That is show that if $1 \leq p < r < q \leq \infty$, and $f \in L^p(\mathbb{R}) \bigcap L^q(\mathbb{R})$, then $f \in L^r(\mathbb{R})$, and*

$$\|f\|_r \leq \|f\|_p^t \|f\|_q^{1-t}, \quad \textit{where} \quad \frac{1}{r} = \frac{t}{p} + \frac{1-t}{q}.$$

Hint: *If $q < \infty$, use Hölder's inequality with dual exponents p/tr and $q/(1-t)r$.*

You have encountered Hölder's inequality (39) at least in the case $q = q' = 2$, asthe Cauchy-Schwarz inequality.

Exercise 65 *Prove the finite dimensional analogue of Hölder, let $a_i, b_i \in \mathbb{R}$, for $i = 1, \ldots, n$,*

$$\sum_{i=1}^n |a_i b_i| \leq \left(\sum_{i=1}^n |a_i|^p\right)^{1/p} \left(\sum_{i=1}^n |b_i|^{p'}\right)^{1/p'}.$$

The case $n = 2$ and $p = 2 = p'$ is just the geometric fact that $|\cos\theta| \leq 1$.

We can deduce the *triangle inequality* in L^p, which is one of the axioms of a norm, from Hölder's inequality. In this case the inequality goes by the name of Minkowski.

Lemma 5 (Minkowski's Inequality) *If $1 \leq p \leq \infty$, and $f, g \in L^p(\mathbb{R})$, then*

$$\|f+g\|_p \leq \|f\|_p + \|g\|_p.$$

Proof: The result is clear if $p = 1$, or if $f + g = 0$. Otherwise, write

$$|f+g|^p \leq (|f| + |g|)|f+g|^{p-1},$$

and apply Hölder's inequality, observing that if $\frac{1}{p} + \frac{1}{p'} = 1$, then $(p-1)p' = p$, therefore,

$$\begin{aligned} \int |f+g|^p &\leq \int |f||f+g|^{p-1} + \int |g||f+g|^{p-1} \\ &\leq \Big(\int |f|^p\Big)^{1/p}\Big(\int |f+g|^p\Big)^{1/p'} + \Big(\int |g|^p\Big)^{1/p}\Big(\int |f+g|^p\Big)^{1/p'} \\ &= \big(\|f\|_p + \|g\|_p\big)\Big(\int |f+g|^p\Big)^{1/p'}. \end{aligned}$$

The left-hand-side term appears on the right-hand-side, we know is not zero, so we get,

$$\|f+g\|_p = \Big(\int |f+g|^p\Big)^{1-1/p'} \leq \|f\|_p + \|g\|_p.$$

Which is the inequality we wanted to show. $\diamond$

One can have n-summands, or even a "continuum" of summands (an integral!). That is the content of the following inequality,

Lemma 6 (Minkowski's Integral Inequality) *Given a function of two variables, $F(x, y)$, such that for a.e. $y \in \mathbb{R}$ the function $F_y(x) = F(y, x)$ belongs to $L^p(\mathbb{R}, dx)$, and the function $G(y) = \|F_y\|_p$ is in $L^1(\mathbb{R})$, then the function $F_x(y) = F(x, y)$ is in $L^1(\mathbb{R}, dy)$ for a.e. $x \in \mathbb{R}$, and the function $H(x) = \int_{\mathbb{R}} F(x, y)\, dy$ is in $L^p(\mathbb{R}, dx)$, moreover,*

$$\|\int_{\mathbb{R}} F(x, y)\, dy\|_{L^p(dx)} \leq \int_{\mathbb{R}} \|F(x, y)\|_{L^p(dx)} dy.$$

You will prove this inequality in the real analysis course. Just keep in mind that it is the triangle inequality for integrals instead of sums.

Corollary 4 (Young's Inequality) *If $f \in L^p(\mathbb{R})$, and $g \in L^q(\mathbb{R})$, then $f * g \in L^r(\mathbb{R})$, where $\frac{1}{r} + 1 = \frac{1}{p} + \frac{1}{q}$, and*

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

Exercise 66 *Prove Young's Inequality by first checking directly that for fixed $g \in L^p$,*

$$\|f * g\|_p \leq \|g\|_p \|f\|_1,$$

and

$$\|f * g\|_\infty \leq \|g\|_p \|f\|_{p'},$$

where $\frac{1}{p} + \frac{1}{p'} = 1$. Then use the Riesz-Thorin Interpolation Theorem. (The second inequality is a direct application of Hölder's inequality. The first one is an easy consequence of Minkowski's Integral Inequality.)

6 Haar functions and the Hilbert transform

6.1 Haar multipliers and the Hilbert transform

We will see that averages over random dyadic grids of appropriate Haar multipliers will give back the Hilbert transform.

A *Haar multiplier* is an operator defined formally by

$$Tf(x) = \sum_{I\in\mathcal{D}} w_I(x)\langle f, h_I\rangle h_I(x);$$

where the symbol $w_I(x)$ is a function of both the space and the "frequency" variables (x, I).

Example 21 *The simplest examples correspond to $w_I(x) = w_I$ and $w_I(x) = w(x)$. In the first case we encounter* constant Haar multipliers*; and in the second multiplication by $w(x)$.*

This is completely analogous to *pseudodifferential operators* where the Haar system has been replaced by the trigonometric functions,

$$\Phi f(x) = \int_R a(x,\xi)\hat{f}(\xi)e^{ix\xi}d\xi,$$

the *symbol* here is $a(x,\xi)$. In both cases one would like to identify those symbols for which the corresponding operators are bounded in, for example, $L^p(\mathbb{R})$.

Example 22 *When the symbol of the pseudodifferential operator is a polynomial in $2\pi i\xi$, $a(x,\xi) = P(2\pi i\xi)$, then Φf coincides with the differential operator $P(D)f$, by the time/frequency dictionary.*

Exercise 67 *What would the pseudodifferential operators with symbol $a(x,\xi) = m(\xi)$ formally correspond to? How about the symbol $a(x,\xi) = a(x)$?*

Exercise 68 *Let T_α be a constant Haar multiplier,*

$$T_\alpha f(x) = \sum_{I\in\mathcal{D}} \alpha_I\langle f, h_I\rangle h_I(x).$$

Show that T is bounded in $L^2(\mathbb{R})$ if and only if $\{\alpha_I\}_{I\in\mathcal{D}}$ is a bounded sequence. (Test action of T on the Haar functions for the necessity. For the sufficiency use Plancherel.) The same result holds for $L^p(\mathbb{R})$, the necessity can be checked testing on Haar functions, and for the sufficiency we can now use the norm equivalence given by the dyadic square function (27) instead of Plancherel.

Constant Haar multipliers are considered models for singular integral operators. In particular, the martingale transform, given by choices of sign σ, introduced in Section 3.4.1,

$$T_\sigma f(x) = \sum_{I\in\mathcal{D}} \sigma_I \langle f, h_I\rangle h_I(x), \qquad \sigma_I = \pm 1.$$

Heuristically it is expected that if certain estimates can be done uniformly on σ for this family, then the same estimates will hold for the Hilbert transform. That has been the driving force behind the work of Nazarov, Treil and Volberg, see [NTV]. But the passage from the martingale transform to the Hilbert transform is not at all obvious, in fact, sometimes it does not seem to work. Very recently, Stephanie Petermichl, a student of Volberg, showed in her PhD Thesis [Pet1], that certain averages over translated and dilated dyadic grids of the operator

$$H_{\mathcal{D}} f(x) = \sum_{I\in\mathcal{D}} \langle f, h_I\rangle \big[h_{I_r}(x) - h_{I_l}(x)\big],$$

produce a non-zero operator which has the following properties:

- It commutes with all translations and dilations.
- It is antisymmetric.

It turns out that the only operators with such properties are constant multiples of the Hilbert transform, therefore they must coincide, rendering now the heuristic very precise, see [Pet2]. Similar results have shown to hold for the *Riesz transforms*, the higher dimensional analogues of the Hilbert transform, see [PV].

7 References

- **[B]** G. Beylkin, *On the representation of operators in bases of compactly supported wavelets.* SIAM J. Numer. Anal. Vol 6, No. 6, pp. 1716-1740 (1992).
- **[BCR]** G. Beylkin, R. Coifman, V. Rokhlin, *Fast wavelet transform and numerical algorithms.* Comm. Pure Appl. Math., **44**, pp. 141-183 (1991).
- **[Cal]** A. Calderón, *Intermediate spaces and interpolation, the complex method.* Studia Math. 24, pp. 113-190 (1964).
- **[Car]** L. Carleson, *On convergence and growth of partial sums of Fourier series.* Acta Math., 116, pp. 135-157 (1966).
- **[CM]** R. Coifman and Y. Meyer, *Remarques sur l'analyse de Fourier à fenêtre.* C. R. Acad. Sci. Paris, Série I, 312, pp. 259-261 (1991).
- **[Dau]** I. Daubechies, *Ten lectures on Wavelets.* CBMS 61, SIAM 1992.
- **[Fr]** M. Frazier, *An introduction to Wavelets through linear algebra.* Springer-Verlag, New York, NY, 1999.
- **[Gab]** D. Gabor, *Theory of Communication.* J. Inst. Electr. Eng., London, 93 (III), pp. 429-457 (1946).
- **[Gr]** L. Grafakos, *Classical and Modern Fourier Analysis.* Prentice Hall 2003.
- **[Ha]** A. Haar, *Zur Theory der orthogonalen Funktionensysteme.* Math. Annal., 69, pp. 331-371 (1910).
- **[HW]** E. Hernandez, G. Weiss, *A first course on wavelets.* CRC Press, Boca Raton, FL, 1996.
- **[Krantz]** S. Krantz, *A Panorama of Harmonic Analysis*, The Carus Mathematical Monographs 27, AMS 1999.
- **[Mall89]** S. Mallat, *Multiresolution approximations and wavelet orthonormal bases for* $L^2(\mathbb{R})$. Trans. AMS, 315, pp. 69-87 (1989).
- **[Mall98]** S. Mallat, *A wavelet tour of signal processing.* Academic Press, San Diego, CA, 1998.
- **[Ma]** H. Malvar, *Lapped transforms for efficient transform /subband coding.* IEEE Trans. Acoustics Speech SIgnal and Processing, 38, pp. 969-978 (1990).
- **[NTV]** F. Nazarov, S. Treil, A. Volberg, *The Bellman function and two weight inequalities for Haar multipliers.* J. Amer. Math. Soc. **12** (1999) n.4, 909–928.

- **[Pet]** S. Petermichl, *Dyadic shifts and a logarithmic estimate for Hankel operators with matrix symbol,* C.R. Acad. Sci. Paris, Sér I Math **330** (2000), 455-460
- **[PV]** S. Petermichl, A. Volberg, *Heating of the Beurling operator: weakly quasiregular maps on the plane are quasiregular,* Duke Math. J. **112** n.2 (2002), 281-305.
- **[Pi]** M. A. Pinsky, *Introduction to Fourier Analysis and Wavelets* The Brooks/Cole Series in Advanced Mathematics, Paul J. Sally, Jr. Editor. Brooks Cole 2001.
- **[RS]** M. Reed, B. Simon, *Methods of modern mathematical physics. II: Fourier analysis, self-adjointness.* Academic Press 1975.
- **[Sha]** C. E. Shannon, *Communications in the presence of noise.* Proc. of the Ins. of Radio Eng., 37, pp. 10-21 (1949).
- **[StSh]** E. Stein, R. Shakarchi, *Fourier Analysis: an introduction.* Princeton Lectures in Analysis I, Princeton Univ. Press 2003.
- **[StWe]** E. Stein, G. Weiss, *Introduction to Fourier analysis on euclidean spaces.* Princeton Univ. Press 1971.
- **[SN]** G. Strang, T. Nguyen, *Wavelets and Filter Banks.* Wellesley-Cambridge Press 1996.
- **[Wick]** V. Wickerhauser, *Adapted wavelet analysis from theory to software.* A. K. Peters, Wellesley, 1994.