Amos Ron
Lectures Notes, CS515
15January02

CS515 Lecture notes

Amos Ron

Part 1. Introduction: Representation

(1:) three examples: lena (four parts) music (48 decompositions) image compression (eeg)

functions: what kind of functions (highly oscillatory, or images). what kind of questions (e.g., the question of: find local minima or local maxima is much less interesting here)

(2:) linear functionals: definition
examples: point evaluation
derivative evaluation
in general: we will assume that the linear functional is given as a function
examples: local averaging, local differencing
(3:) definition of the analysis map (=: the decomposition map)

$$\Lambda^* : f \mapsto \{\lambda_i f\}_{i\in I}.$$

The set I is some index set that we use in order to index our linear functionals. For example, the positive integers $\mathbb{N}$, the integers $\mathbb{Z}$ etc.

stress: decomposing the function with a well-organized collection of linear functionals.

the simplest example: the (regular) sampling transform
Definition: a *signal* is the image of a function under the regular sampling transform
Note: there is a natural way to order the samples (in this case the ordering coincides with the coordinates of the original function).

Note: the domain (time) of the function is, for regular sampling, essentially the same of the domain of the transformed function (i.e., the signal)

Organization of the decompositions: we always need to have some good way to organize the linear functionals of the decomposition map. This means that the linear functionals are selected from a well-defined closely related family (e.g., in the case of sampling all the linear functionals are point-evaluations).

(4:) Basic assumption: time is invariant.

Thus: if we λ_g is a linear functional, all linear functionals induced by **translations** of g should be in.

Translation:

$$E^t(f)(u) := f(u-t).$$

this means that we *convolve*.

definition of **convolution**:

$$(f * g)(t) := \int_{\mathbb{R}} f(u)g(t-u)\,\mathrm{d}u.$$

Note that we are 'flipping' the function g before we start to compute the inner products between f and the translates of g. One explanation for that is the desire to get nice properties for the binary operation $f * g$:

properties: commutative $f * g = g * f$, associative $(f * g) * h = f * (g * h)$, distributive $f * (g + h) = (f * g) + (f * h)$.

example: $B_1 * B_1 = B_2$.

---

exponential functions:

cos, sin, exp, the notion of frequency

modulation as a way to measure changes in a function

leads to the fourier transform and to fourier analysis

---

dilation: given a dilation parameter $a > 0$, the a-dilation of f is the function

$$t \mapsto f(at).$$

Part 2. Introduction: Fourier series and orthonormal systems.

the basic assumption: all the linear functionals are exponentials.

how many exponentials?

two different theories: fourier series, the function is supported on an interval.

we choose the interval to be $[-\pi, \pi]$ for convenience.

fourier transform, or fourier integral, the functions are supported on $\mathbb{R}$.

Fourier series.

what kind of functions we would like to decompose?

the L_2-space: definition: a function f is in L_2 if

$$\|f\|_{L_2} := (\int_{-\pi}^{\pi} |f(t)|^2\,dt)^{1/2}$$

is finite.

Note: the inner product

$$\langle f, g\rangle := \int_{-\pi}^{\pi} f(t)\overline{g(t)}\,dt$$

is connected to the norm via the relation

$$\|f\|^2 = \langle f, f \rangle.$$

This means that L_2 is *an inner product space.*

Examples: the (restriction to $[-\pi, \pi]$ of) the function $|t|^{-1/2}$ is not in L_2. However, $|t|^{-2/3}$ is in L_2 (the actual definition of these functions at the origin is immaterial).

Definition: the Fourier series. Let $e_{i\omega}$ be the periodic exponential with frequency ω, i.e.,

$$e_{i\omega} : t \mapsto e^{i\omega t} = \cos(wt) + i \sin(wt).$$

The linear functionals in the decomposition map of the Fourier series are the *exponentials with integer frequencies* i.e.,

$$\Lambda := (e_{in})_{n=-\infty}^{\infty}.$$

Explicitly, we denote $\widehat{f}(n) := \langle f, e_{in} \rangle$, i.e.,

$$\widehat{f}(n) := \int_{-\pi}^{\pi} f(t) e^{-int}\, dt.$$

We then consider (naturally) $\widehat{f}$ as a function (i.e., sequence) defined on the integers $\mathbb{Z}$.

There is a sequence space, ℓ_2, which is analogous to the L_2-space: A sequence $x : \mathbb{Z} \to \mathbb{C}$ belongs to ℓ_2 (or the sequence is 'square-summable') if

$$\|x\| := \left(\sum_{n=-\infty}^{\infty} |x(n)|^2\right)^{1/2} < \infty.$$

Properties of the analysis map of the Fourier series.

orthogonality: $\langle e_{in}, e_{im} \rangle = 0$, for any two different integers $n.m$. Moreover, $\langle e_{in}, e_{in} \rangle = 2\pi$ (check!). (Thus, had we chosen to 'normalize' each exponential, by dividing it by $\sqrt{2\pi}$ we would have obtained an *orthonormal* system, i.e., an orthogonal system with the norm of each element being unit; because we do not normalize the exponentials, the factor 2π occurs in all the formulas below).

An important consequence of the above orthogonality is the **Bessel inequality**:

$$\|\widehat{f}\| \le \sqrt{2\pi}\, \|f\|, \quad \forall f \in L_2,$$

i.e.,

$$\left(\sum_{n=-\infty}^{\infty} |\widehat{f}(n)|^2\right)^{1/2} \le \sqrt{2\pi}\, \|f\|_{L_2}.$$

In particular, the map $f \mapsto \widehat{f}$ maps L_2 into ℓ_2.

completeness: This is a highly non-trivial property of the exponential system $(e_{in})_{n\in\mathbb{Z}}$ and is known as the Fischer-Riesz Theorem. There are several different ways to express this property. The most convenient one is to say that 'the map $f \mapsto \widehat{f}$ is one-to-one' i.e., 'the only function f in L_2 that satisfies $\widehat{f}(n) = 0$, all n, is the zero function.'

The fact that the exponential system is complete and (almost) orthonormal implies several very important properties.

First and foremost is the *perfect reconstruction property*:

Theorem. *For every* $f \in L_2$,

$$f = \frac{1}{2\pi} \sum_{n-\infty}^{\infty} \widehat{f}(n) e_{in} = \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} \langle f, e_{in} \rangle e_{in}.$$

The second property is *Parseval's identity*: for every $f \in L_2$,

$$\|f\| = \frac{1}{\sqrt{2\pi}} \|\widehat{f}\|.$$

Another, seemingly stronger, version of Parseval's identity is:

$$\langle f, g \rangle = \frac{1}{2\pi} \langle \widehat{f}, \widehat{g} \rangle,$$

i.e.,

$$\int_{\mathbb{R}} f(t)\overline{g(t)}\, dt = \frac{1}{2\pi} \sum_{k \in \mathbb{Z}} \widehat{f}(k)\overline{\widehat{g}(k)}.$$

Note: Parseval's identity is *equivalent* to the perfect reconstruction property. None of the two implies the orthonormality of the system (of linear functionals) that we use. Later on, we will see systems of linear functionals that give us perfect reconstruction without being orthonormal (or orthogonal).

Connection between smoothness of f and the decay of its Fourier coefficients.

Theorem. *Let k be a positive integer. Suppose that all the derivatives up to order k of f (as a 2π-periodic function) exist and lie in L_2. Then the sequence $n \mapsto n^k \widehat{f}(n)$ lies in ℓ_2. The converse is also true.*

Example: take the function f which is 0 on $[-\pi, 0)$ and is 1 on $[0, \pi)$ (note that the function has two 'bad' points: the obvious one is at 0, the less obvious one is at π; recall that we think of the function as 2π-periodic hence identify π with $-\pi$.) This function does not satisfy the above theorem for any value of k (other than $k = 0$), hence we are granted that the sequence $n \mapsto n\widehat{f}(n)$ is not square-summable. This implies, for example, that we *cannot* have an inequality of the form

$$|\widehat{f}(n)| \le c\,|n|^{-1.5-\varepsilon}, \quad \forall n \in \mathbb{Z},$$

for some $c, \varepsilon > 0$ (why?). Compute $\widehat{f}$ and find the exact rate of decay of it. □

Discussion. The above exemplifies the main shortcoming of the Fourier series: it represents the original function f on a new domain (as the sequence $\widehat{f}$) in a way which is not local in time. The function f in the above example has a bad behavior at two points, and is very nice elsewhere. The decomposed function $\widehat{f}$ reacts to the bad points on the time domain by decaying very slowly (as it must do, in view of the above theorem). While one can immediately conclude from that decay that the original f cannot have a first order derivative in L_2, there is (at least in essence) no way to tell where the bad points of f are.

Part 3. Introduction: Fourier transform (L_2 theory)

In many (but not all) regards, the Fourier transform is the extension of the Fourier series theory from periodic functions to functions defined on the entire real line.

This will be the first and last time that we use a non-countable number of linear functionals (i.e., there are so many linear functionals that there is no way to index them by the integers).

The L_2-space. We now assume that our functions are defined on the entire line. Then our inner product is

$$\langle f, g\rangle := \int_{\mathbb{R}} f(t)\overline{g(t)}\, dt,$$

our norm is

$$\|f\|_{L_2} := \Big(\int_{\mathbb{R}} |f(t)|^2\, dt\Big)^{1/2},$$

and the space L_2 is again the space of all functions whose above L_2-norm is finite. Note that the functions in L_2 should not be too bad around any point, and also should decay somewhat at $\pm\infty$.

Examples. The function $t \mapsto 1/t$ is not in L_2 because it behaves too badly at the origin. The function $t \mapsto t^{-1/3}$ is not in L_2 because it does not decay fast enough as t approaches $\pm\infty$. The function $t \mapsto 1/\sqrt{|t|}$ is 'too bad' both at 0 and at $\pm\infty$. How about

$$t \mapsto \frac{1}{|t| + \sqrt{t}}?$$

The Fourier transform $f \mapsto \widehat{f}$ is the decomposition map that employs *all* the exponentials

$$\Lambda := (e_{i\omega})_{\omega\in\mathbb{R}}.$$

Thus,

$$\widehat{f}(\omega) := \langle f, e_{i\omega}\rangle = \int_{\mathbb{R}} f(t)e^{-i\omega t}\, dt.$$

The are a few technical difficulties here that arise from the fact that the exponentials (viewed as functions defined on the entire line) do not belong to the L_2-space. For example, the notion of orthogonality is not very meaningful since it is hard to make sense of the inner product of two exponentials. Also, the above definition of the Fourier transform makes sense only for 'nice enough' functions f in L_2.

We brush away all these difficulties, partly since we really are interested in analysing functions f of *compact support* (and for those the theory is simpler).

We regard $\widehat{f}$ as a function defined also on $\mathbb{R}$, i.e.,

$$\widehat{f} : \mathbb{R} \to \mathbb{C} : \omega \mapsto \widehat{f}(\omega).$$

However, the fact that the domains of f and $\widehat{f}$ seem to be the same, is misleading (it is mathematical accident: recall that in the periodic case, f is defined on $[-\pi, \pi]$ while $\widehat{f}$ is defined on $\mathbb{Z}$). Many books make the distinction formal by denoting the domain of $\widehat{f}$ by $\widehat{\mathbb{R}}$, although of course $\widehat{\mathbb{R}}$ is still the real line. We will distinguish between the two domains by referring to one of them as the **time domain** and the other as the **frequency domain**.

The core of Fourier analysis is the fact that f and $\widehat{f}$, while both defined on the same real line, exhibit completely different behavior. What may be apparent from an inspection of f (e.g., a jump discontinuity) may be very hard to observe by looking at $\widehat{f}$, and vice versa. To a large degree, we would like to be able to look at a function *simultaneously in both domains.*

Example: Music. Music is an excellent example of the combined meaning of time and frequency. We may regard each note of an instrument as representing one particular frequency (it does not matter for the present discussion whether this is completely true). So, the most basic info about music (one instrument, say), is to know which note was played and when. The time representation of music answers the question 'when': the music was played during the time that its time representation was non-zero (the time representation, in essence, records the amplitude of the music at each particular time). The frequency representation (i.e., $\widehat{f}$) answers the question 'which': it tells us what notes were active during the entire time that the music was played. Neither of the two is satisfactory. We will be looking soon for methods that allows a simultaneous representation of a function on a combined 'time-frequency' domain. We need first the Fourier transform, since it defines for us a new domain for inspecting a function the *frequency domain.*

Properties of the Fourier transform (i.e., properties of the above exponential set Λ).

Completeness. The Fourier transform is one-to-one on L_2, i.e., the only function in L_2 that satisfies $\langle f, e_{i\omega}\rangle = 0$ for each $\omega \in \mathbb{R}$ is the zero function.

Parseval identity. For every $f \in L_2$,

$$\|f\| = \frac{1}{\sqrt{2\pi}}\|\widehat{f}\|.$$

Again, this identity leads to an analogous result on the corresponding inner products:

$$\langle f, g\rangle = \frac{1}{2\pi}\langle \widehat{f}, \widehat{g}\rangle,$$

i.e.,

$$\int_{\mathbb{R}} f(t)\overline{g(t)}\,dt = \frac{1}{2\pi}\int_{\mathbb{R}} \widehat{f}(t)\overline{\widehat{g}(t)}\,dt.$$

Perfect reconstruction.

$$f(t) = \frac{1}{2\pi}\langle \widehat{f}, e_{-it}\rangle = \frac{1}{2\pi}\int_{\mathbb{R}} \widehat{f}(\omega)e^{i\omega t}\,d\omega.$$

The next three properties are easy to prove (try!)

Connection between translation and modulation. For every $t \in \mathbb{R}$,

$$\widehat{e_{it} f} = E^t \widehat{f}, \quad \widehat{E^t f} = e^{-it} \widehat{f}.$$

I.e., translation on the time domain is converted to modulation of the frequency domain. (The fact that modulation of the time domain is converted to translation of the frequency domain must then follow, since applying twice the Fourier transform brings us back, almost exactly, to the original function; see the Perfect Reconstruction property).

connection between convolution and multiplication. This is among the most remarkable and the most powerful properties of the Fourier transform:

$$\widehat{f * g} = \widehat{f}\widehat{g}. \tag{1}$$

Example. Let B_1 be the B-spline of order 1. It is relatively easy to compute its Fourier transform

$$\widehat{B_1}(\omega) = \frac{1 - e^{-i\omega}}{i\omega}.$$

Higher order B-splines are defined by repeated convolutions:

$$B_k := B_{k-1} * B_1.$$

It is non-trivial to compute B_k (it is, btw, a piecewise-polynomial supported on $[0, k]$). The property (1) implies, almost immediately, that

$$\widehat{B_k}(\omega) = \left(\frac{1 - e^{-i\omega}}{i\omega}\right)^k.$$

connection between dilation and dilation. It is useful to define dilation in a normalized way: if $a > 0$, then

$$(\mathcal{D}_a f)(t) := \sqrt{a} f(at).$$

(In this way, $\|f\| = \|\mathcal{D}_a f\|$.)

Then

$$\widehat{\mathcal{D}_a f} = \mathcal{D}_{1/a} \widehat{f}.$$

(Thus, dilation by a on the time domain is converted to dilation by $1/a$ on the frequency domain: 'stretching' on the time domain becomes 'squeezing' on the frequency domain.)

connection between differentiation and multiplication by a polynomial; connection between smoothness of f and decay of $\widehat{f}$.

Let $()$ be the linear function $\omega \mapsto \omega$. Then

$$\widehat{f'} = ()\widehat{f}.$$

It follows:

Theorem. *Let k be a positive integer and let $f \in L_2$. Then the derivatives $f', f'', \ldots, f^{(k)}$ all exist and lie in L_2 if and only if the function $()^k \widehat{f} : \omega \mapsto w^k \widehat{f}(\omega)$ lies in L_2.*

Example. We take the function B_k:= the B-spline of order k. While we do not know much (yet) about this function in the time domain, we already know that

$$\widehat{B_k}(\omega) = \left(\frac{1 - e^{-i\omega}}{i\omega}\right)^k.$$

We can bound

$$|\widehat{B_k}(\omega)| \leq 2^k |\omega|^{-k}.$$

Thus,

$$|\omega^{k-1} \widehat{B_k}(\omega)| \leq 2^k |\omega|^{-1}.$$

This implies that $()^{k-1}\widehat{B_k} \in L_2$, hence that all the derivatives of B_k up to order $k-1$ exist and are in L_2. □

Part 4. Time-frequency localization and WH systems

We would like to construct systems that:

(1) Perform a good time-frequency localization. In principle, this means that the functions in the system are *local in time* (e.g., compactly supported, or decay very fast at ∞), and are also very smooth (since this corresponds to good decay of their Fourier transform).

(2) Are good in the sense of the section on 'Good Systems'. This means that once we applied the decomposition operator

$$\Lambda^* : f \mapsto \langle f, \lambda \rangle_{\lambda \in \Lambda},$$

we have a 'good' way to reconstruct f. For example, a very good system would allow us a *perfect reconstruction*:

$$f = \sum_{\lambda \in \Lambda} \langle f, \lambda \rangle \lambda.$$

(2) Are augmented by a fast algorithm that allows us to do painlessly the decomposition and reconstruction.

When attempting to perform good time-frequency localization, we must have certain priorities in mind: there is a subtle balance between the ability to be 'very local' in time, and the ability to be very local in frequency.

Most of the prevailing constructions start with a window function g (or several window functions) and associate with it its set of **shifts** i.e., *integer* translations:

$$E(g) := \{E^k g : \ k \in \mathbb{Z}\}, \qquad E^k g : t \mapsto g(t-k).$$

Assuming that g is 'concentrated' around the origin, we may associate the function $E^k g$ in the system $E(g)$ with the point $t = k$ in the time domain.

In order to complete the construction of the system Λ, we need to choose between the following two options:

Option 1: we aim at having elements in Λ that identify very local features in the time domain. If this is our goal, then we need to have in Λ functions of smaller and smaller supports. This can be achieved by applying *dilations* to $E(g)$, and it leads to the notion of *wavelets*. The sacrifice here is in the frequency domain: using such a system, our ability to distinguish between frequencies will deteriorate as the frequency gets higher.

Option 2: we would like to 'tile' the frequency domain in a similar way to that of the time domain. I.e., given the window g, we would like that our system will include all the functions whose Fourier transforms is of the form $E^j \widehat{g}$, with j varies over either the integers of some fixed scale of the integers. This approach leads to the notion of *Weyl-Heisenberg systems*. Since translating the Fourier transform is equivalent to modulating the original function, we are led to constructing here a system of the form

$$\Lambda = \{e_{ij} E^k g : \ k \in \mathbb{Z}, \ j \in 2\pi\mathbb{Z}\}.$$

I.e., a typical function in the system is of the form

$$g_{j,k} : t \mapsto e^{ijt} g(t - k).$$

If $\widehat{g}$ is 'concentrated' around the origin (in the frequency domain), then $\widehat{g_{j,k}}$ is concentrated around $2\pi j$. This means that, roughly speaking, that the inner product

$$\langle f, g_{j,k} \rangle$$

'tells us' about the behaviour of f at time $t = k$ and frequency $\omega = 2\pi j$. (The fact that we use the lattice $2\pi\mathbb{Z}$ on the frequency domain is not essential, and for this reason we do not justify that choice; however, see the theorem below).

Example. Let B_1 be the B-spline of order 1. Then the Weyl-Heisenberg (WH) system

$$\{e_j E^k B_1 : \ k \in \mathbb{Z}, \ j \in 2\pi\mathbb{Z}\}$$

is known as the (discretized) *windowed Fourier transform*. It is a complete orthonormal system (and therefore has the 'perfect reconstruction' property. It is local in time (although its elements cannot 'zoom on' very local features, which is a drawback of all WH systems), but its localness in frequency of very bad (why? look at $\widehat{B_1}$). □

The attempt to construct WH systems with better frequency localization has to deal first the following theoretical barriers. The third part of the theorem in known as the *Balian-Low Theorem*.

Theorem 2. *Let $g \in L_2$, and let Λ be the*

$$\Lambda := \{e_j E^k g : \ k \in \mathbb{Z}, j \in h\mathbb{Z}\}.$$

Then:

(i) If $h < 2\pi$ (=oversampling) the system Λ is dependent, i.e., one of the elements in the system can be represented by the others. In particular, Λ cannot be orthonormal in this case.

(ii) If $h > 2\pi$ (=undersampling) the system Λ is not complete.

(iii) If $h = 2\pi$, and if Λ is known to be complete and orthonormal, then either $g' \notin L_2$ (and then the system has very poor frequency localization), or $\widehat{g}' \notin L_2$ (and then the system has very poor time localization).

One may attempt to conclude from the above that there is no way to construct good WH systems. That is not the case however. First, there is a genuine trick that alters a bit the definition of a WH system. The systems constructed in this twisted manner are known as *Wilson bases* and they escape the curse of the Balian-Low theorem: there are smooth compactly supported Wilson bases which are orthonormal and complete. The famous construction of Wilson bases is due to Daubechies-Jaffard-Journé (1992).

Simpler than that: it is very easy to construct WH systems which satisfy the complete reconstruction property, and have excellent time-frequency localization property as well (they are not orthonormal however). The first such construction is due to Daubechies-Grossman-Meyer (1986).

Theorem 3. *Let g be a function supported in the interval $[0, 1/h]$, for some positive $h < 1$. Then the system*

$$\Lambda = \{e_j E^k g : \ j \in \mathbb{Z}, \ k \in 2\pi h\mathbb{Z}\}$$

is a complete orthonormal system, if and only if

$$\sum_{k \in \mathbb{Z}} |g|^2(\cdot + k) = 1.$$

There are many compactly supported univariate functions whose shifts sum up to the constant 1. For example, this is true for each B-spline B_m. Thus, we can take g to be the square root of the B-spline B_m. Since B_m is supported in the interval $[0, m]$ we can choose $h = 1/m$ for this case.

Part 5: Wavelets and MRA

Wavelet systems are created from the shift-invariant system $E(g)$ by applying *dilations.* The most standard dilations are in powers of 2. Such wavelet systems are sometime referred to as ‘dyadic wavelets’.

Definition: a (dyadic) wavelet system. Let Ψ be a finite collection of a functions in $L_2(\mathbb{R})$. The wavelet system generated by Ψ is the collection of functions

$$W_\Psi := \{\mathcal{D}_{2^j} E^k \psi : \ j, k \in \mathbb{Z}, \ \psi \in \Psi\}.$$

The functions in Ψ are called **mother wavelets**. We index the wavelet by ψ, j and k. Thus,

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

Note: $\psi_{j,k}$ is obtained from $\psi_{j,0}$ by translation. The translation is not by k, but by $k/2^j$. Thus, our shifts become denser as $j \to \infty$ and sparser as $j \to -\infty$. This completely agrees with the fact that positive dilation 'squeezes' the function, while negative dilation 'stretches' the function.

Proposition. *Let W_ψ be a wavelet system generated by a single mother wavelet ψ. Then W_ψ is orthonormal if and only if the following condition is valid for every $k \in \mathbb{Z}$ and every $j \geq 0$:*

$$\langle \psi_{0,0}, \psi_{j,k} \rangle = \begin{cases} 1, & j = k = 0, \\ 0, & \textit{otherwise}. \end{cases} \tag{4}$$

Proof: One implication is trivial, i.e., if the system is orthonormal then the above three conditions should definitely hold: they are a part of the *definition* of an orthonormal system.

In order to prove the converse, we assume that (4) hold. We note that both dilation and translation are unitary operators, that is for every $t \in \mathbb{R}$ and every non-zero a and every $f, g \in L_2$

$$\langle E^t f, E^t g \rangle = \langle \mathcal{D}_a f, \mathcal{D}_a g \rangle = \langle f, g \rangle.$$

We also note that

$$\mathcal{D}_a E^t = E^{t/a} \mathcal{D}_a.$$

Thus, first,

$$\langle \psi_{j,k}, \psi_{j,k} \rangle = \langle \mathcal{D}_{2^j} E^k \psi, \mathcal{D}_{2^j} E^k \psi \rangle = \langle \psi, \psi \rangle = 1.$$

Second, in computing $\langle \psi_{j,k}, \psi_{j',k'} \rangle$ we may assume that $j \leq j'$ (since the inner product is symmetric, up to conjugation). Thus,

$$\langle \psi_{j,k}, \psi_{j',k'} \rangle = \langle \mathcal{D}_{2^j} E^k \psi, \mathcal{D}_{2^{j'}} E^{k'} \psi \rangle = \langle E^k \psi, \mathcal{D}_{2^{j'-j}} E^{k'} \psi \rangle =$$

$$\langle \psi, E^{-k} \mathcal{D}_{2^{j'-j}} E^{k'} \psi \rangle = \langle \psi, \mathcal{D}_{2^{j'-j}} E^{k' - 2^{j'-j} k} \psi \rangle = 0,$$

with the last equality by assumption (4), since $\mathcal{D}_{2^{j'-j}} E^{k' - 2^{j'-j} k} \psi = \psi_{j'-j, k' - 2^{j'-j} k}$. □

We organise the wavelets in W_Ψ in *scales* or *layers* of *levels*. The jth scale, W_j of the system consists of the $\mathbb{Z}/2^j$-shifts of the dilated functions $\psi_{j,0}$.

$$W_j := \{\psi_{j,k} : \ \psi \in \Psi, \ k \in \mathbb{Z}\}.$$

Let's see a few examples.

Example: The Haar wavelet system. We choose Ψ to consist only of one function, and take this function to be the Haar function H. Then, one can show that W_H is orthonormal (easy) and complete (a bit harder). This is the archetypal wavelet system. □

The Haar system gives us a good insight to the way wavelets organize the time domain: each scale W_j covers completely the time line $\mathbb{R}$. As $j \to \infty$, the 'tiles' becomes smaller (hence we need more of them). As $j \to -\infty$ the tiles becomes larger...

It is hard to envision the frequency decomposition of the wavelets by looking at the Haar system. The performance of Haar on the frequency domain is so poor, that it gives there a very blurred picture of 'what should have happened'.

It is useful therefore to go to the other extreme and to look at the wavelet system which has the ideal frequency localization (and which is very poorly localized on the time domain).

The Shannon wavelet system. We take Ψ to consist again of a single mother wavelet. We choose this wavelet by defining its Fourier transform:

$$\widehat{\psi}(\omega) := \begin{cases} 1, & \pi \le |\omega| \le 2\pi, \\ 0, & \text{otherwise.} \end{cases}$$

(one of the homework problems will ask you to compute this function explicitly.) Again, it is easy to prove that the Shannon wavelet system is orthonormal (it is also complete). It is also easy to understand that the wavelets of this system have poor time localization (why?). However, the point of this example is to see how the wavelets tile here the frequency domain. The is *the ideal frequency tiling* (for wavelets).

The Fourier transform of $\psi_{0,0}$ is given explicitly. Recall that a dilation on the time domain is transformed to the opposite dilation on the frequency domain. Thus, the Fourier transform of the 'squeezed' wavelet $\psi_{1,0}$ is the support function of the 'stretched' domain:

$$\widehat{\psi}_{1,0}(\omega) := \frac{1}{\sqrt{2}} \begin{cases} 1, & 2\pi \le |\omega| \le 4\pi, \\ 0, & \text{otherwise.} \end{cases}$$

let's accept for the time being the concept that we would like to keep the supports of the various $\widehat{\psi_{j,0}}$ as separated as possible. The Shannon wavelet is ideal in this regard, since that support of $\{\widehat{\psi_{j,0}}\}_{j=-\infty}^{\infty}$ tile the frequency domain.

The important observation here that we need $\widehat{\psi_{0,0}}$ not only to decay fast at ∞, but we also need it to have a high order zero at the origin: otherwise, we get substantial overlaps of the supports of $\widehat{\psi_{j,0}}$ for *negative* j. We amplify that in the next discussion.

Thus, we think about the wavelet as 'band pass filters' which means that their frequency content should concentrate in a domain that is 'between' 0 at ∞. □

Discussion. By now, we know of a few ways to judge whether a given system is good or not. One criterion is the ability to invert the decomposition by a 'good reconstruction'. As said, complete orthonormal systems enjoy the perfect reconstruction property. Thus, in this regard the Haar system is 'perfect'.

Another criterion is the time-frequency localization. In terms of time localization the Haar function is perfect. Its frequency localization is poor. There are now *two* different criteria that should be satisfied when judging the frequency localization of the wavelet ψ.

How to judge the frequency localization of a given mother wavelet ψ? The first is by looking at the decay rate of $\widehat{\psi}$ at $\pm\infty$, or equivalently, at the smoothness of ψ. The Haar wavelet already fails this initial test. But it also fails another 'frequency localization test' which is specific to wavelets only: its Fourier transform has only a first order zero at the origin. □

The main issue at stake: How to construct a wavelet system that (i) is 'good' (e.g., complete and orthonormal), (ii) performs good time-frequency localization (the mother wavelets are compactly supported, smooth, and their Fourier transform vanishes to a high order at the origin), (iii) can be implemented by a fast algorithm. Sometime, we also add another requirement: (iv) the mother wavelet are symmetric (or anti-symmetric; note that the Haar wavelet is anti-symmetric around the point 1/2.)

The vehicle for constructing wavelets systems is *MultiResolution Analysis* (MRA). It was introduced by Mallat and Meyer in the late 80's. At the heart of MRA is the notion of a *refinable function* (also known as 'scaling function' and 'father wavelet').

Definition 5. Let $\phi \in L_2$ be a given a function. We say that the function ϕ is **refinable** if we can write ϕ as linear combination of the half-shifts of $\mathcal{D}_2\phi$. □

We will alter a bit this definition in the sequel. But, let's start with examples.

Example. The simplest example of a refinable function is $\phi := B_1$, since it is clear that for this ϕ

$$\phi(t) = \phi(2t) + \phi(2t-1). \tag{6}$$

We can write this relation also as

$$\sqrt{2}\phi = \phi_{1,0} + \phi_{1,1}.$$

□

Thus, to say that ϕ is refinable it tantamount to saying that

$$\phi = \sum_{k\in\mathbb{Z}} c(k)\phi_{1,k},$$

for a suitable sequence $(c(k))_{k=-\infty}^{\infty}$. For technical reasons, we normalize this sequence and introduce

$$h(k) := \frac{c(k)}{\sqrt{2}}.$$

For example, in the case of B_1, $h(0) = h(1) = 1/2$, and $h(k) = 0$, for all other values of k.

The above sequence h is at the core of MRA and is usually referred to as the **mask** of the refinable ϕ. In most (but not all) examples of interest, the mask h is *finitely supported* i.e., while being formally defined on all the integers, it assumes non-zero values only at finitely many integers.

It will be convenient thus to define a mask by specifying only the non-zero entries of it. Thus, we could simply define the mask of B_1 by $h(0) = h(1) = 1/2$.

Another useful notation here is the following:

$$V_0(\phi) := \text{ all the linear combination of } E(\phi).$$

The definition is a little vague: the exact definition is that $V_0(\phi)$ is the L_2-closure of the finite linear combinations of the shifts $E(\phi)$ of ϕ. A more down-to-earth definition (which, unfortunately works only in the case ϕ is compactly supported) is to take all linear combinations of $E(\phi)$ (finite or not) that are 'meaningful' and which are in L_2.

It is more important here to pay attention to the nature of $V_0(\phi)$ then to its exact definition. For example, if $\phi = B_1$, then $V_0(\phi)$ consists of piecewise-constants with (possible) integer breakpoints. If $\phi = B_2$, then $V_0(\phi)$ consists of continuous piecewise-linear functions with integer breakpoints.

Set

$$V_1(\phi) := \mathcal{D}_2 V_0(\phi).$$

Thus, $V_1(\phi)$ is obtained by applying dyadic dilation to all the functions in $V_0(\phi)$. It is easy to see that $V_1(\phi)$ is the 'span' of $\phi_{1,k}$, $k \in \mathbb{Z}$, i.e. it is spanned by the *half-shifts* of the dilated ('squeezed') function $\phi_{1,0}$.

Note that, in the above term, the refinability assumption simply says that

$$\phi \in V_1(\phi).$$

Example. Let ϕ be the hat function $B_2(\cdot + 1)$. Then (check!)

$$\phi = \frac{1}{\sqrt{2}}(\frac{1}{2}\phi_{1,-1} + \phi_{1,0} + \frac{1}{2}\phi_{1,1}).$$

This means that the *centered* hat function is refinable with mask $h(-1) = h(+1) = 1/4$, and $h(0) = 1/2$. □

There is yet another, sometimes more convenient, way to understand the notion of refinability: viewing the refinability condition as connecting the Fourier transforms of ϕ and $\mathcal{D}_2\phi$. Apply, for example, the Fourier transform to equation (6). Then (after some simple calculation) we get that (for $\phi := B_1$),

$$\widehat{\phi}(\omega) = \frac{1}{2}(1 + e^{-i\omega/2})\widehat{\phi}(\omega/2).$$

If we set

$$H(\omega) := \frac{1 + e^{-i\omega}}{2},$$

then we can write

$$\widehat{\phi}(\omega) = H(\omega/2)\widehat{\phi}(\omega/2).$$

The most important thing to observe is that $H(\omega)$, as a linear combination of periodic exponentials, is 2π*-period*. In fact, since it is a *finite* linear combination of such exponentials, we name it a **trigonometric polynomial**. Thus, we found that on the Fourier domain the refinability of ϕ leads to the existence of a 2π-periodic function H together with the relation

$$\widehat{\phi} = H(\cdot/2)\widehat{\phi}(\cdot/2).$$

The 2π-periodic function H is nothing but the Fourier series of the mask sequence h:

$$H(\omega) = \sum_{k=-\infty}^{\infty} h(k)e^{-i\omega k}.$$

The function H is also called the mask of the refinable ϕ, or the **symbol** of ϕ.

Example. We previously found that the mask of the centered hat function is defined by $h(-1) = h(1) = 1/4$, $h(0) = 1/2$. Thus, the symbol of this mask is

$$H(\omega) = \frac{e^{-i\omega} + e^{i\omega}}{4} + \frac{1}{2} = \frac{\cos(\omega) + 1}{2} = \cos^2(\omega/2).$$

We can find that out directly, by finding first the Fourier transform of the centered hat function. Recall that the transform of B_1 is

$$\widehat{B_1}(\omega) = \frac{1 - e^{-i\omega}}{i\omega} = e^{-i\omega/2}\frac{e^{i\omega/2} - e^{-i\omega/2}}{i\omega} = e^{-i\omega}\frac{\sin(\omega/2)}{\omega/2}.$$

Thus, the transform of the non-centered hat function B_2 is

$$\widehat{B_2}(\omega) = e^{-i\omega}\big(\frac{\sin(\omega/2)}{\omega/2}\big)^2.$$

Since our function ϕ is $E^{-1}B_2$, we obtain that

$$\widehat{\phi}(\omega) = \big(\frac{\sin(\omega/2)}{\omega/2}\big)^2.$$

Thus, the symbol H is defined by the relation

$$\big(\frac{\sin(\omega/2)}{\omega/2}\big)^2 = H(\omega/2)\big(\frac{\sin(\omega/4)}{\omega/4}\big)^2.$$

Using the correct trigonometric identity (which one?) one gets the exact formula for H.
□

We now make a more convenient definition of refinability:

Definition 7. Let ϕ be in L_2. We say that ϕ is **refinable** if there exists a 2π-periodic function $H(\omega)$ such that

$$(8). \qquad \widehat{\phi}(2\omega) = H(\omega)\widehat{\phi}(\omega)$$

Example. Let ϕ be the function whose Fourier transform is the support function of the interval $[-\pi, \pi]$ (i.e. $\widehat{\phi}$ equals 1 on that interval, and 0 outside that interval). Let H be the (2π-periodic extension) of the support function of the interval $[-\pi/2, \pi/2]$. Then, obviously, ϕ is refinable (according to definition (7)) with mask H. □

There are only a handful of known refinable functions. On the other hand, the second definition of refinability allows one to construct, at least in a formal way, many such functions. Simply, let H be some 2π-periodic function, and define

$$(9) \qquad \widehat{\phi}(\omega) := \prod_{j=1}^{\infty} H(\omega/2^j).$$

Theorem 10. *Let H be a trigonometric polynomial (i.e., a finite linear combination of exponentials), and assume that*

$$H(0) = 1.$$

Then the infinite product (9) converges everywhere.

The theorem is insufficient: it does not tell us that the limit is the Fourier transform of some $\phi \in L_2$. It turns out that this is a much harder problem. Here is a much stronger theorem, in which we assume H to be a CQF:

Definition 11. Let H a 2π-periodic function. We say that H is a CQF (Conjugate Quadrature Filter) if $H(0) = 1$ and

$$|H(\omega)|^2 + |H(\omega + \pi)|^2 = 1.$$

Example. Let

$$H(\omega) = e^{i\omega/2}\cos(\omega/2) = \frac{1 + e^{i\omega}}{2}.$$

Then $|H(\omega)|^2 = \cos^2(\omega/2)$, and it follows then that H satisfies the CQF condition. Note this particular H is the mask of B_1.

Theorem 12. *Let H be a trigonometric polynomial, and assume that H satisfies the CQF condition.*
Then there exists a compactly supported function $\phi \in L_2$ which is refinable with mask H (and whose Fourier transform satisfies (8) for the current H). Moreover, ϕ satisfies one (and only one) of the following two conditions:
(1) The shifts $E(\phi)$ of ϕ are orthonormal

(2) $\widehat{\phi}$ *has a* 2π*-periodic zero, i.e., there exists a point* ω_0 *such* $\widehat{\phi}(\omega_0 + 2\pi m) = 0$, *for every integer* m.

Discussion and example. In general, we would like to conclude that the refinable function whose mask is a CQF has orthonormal shifts. The above theorem simply says that the orthonormality condition of $E(\phi)$ is implied by the CQF condition of H, once we know that $\widehat{\phi}$ does not have 2π-periodic zero. As an example, take

$$H(\omega) := \frac{1 + e^{-3i\omega}}{2}.$$

Then H is a trigonometric polynomial, $H(0) = 1$ and H satisfies the CQF condition (check!). The refinable function is the support function of the interval $[0, 3]$, which obviously does not have orthonormal shifts. Indeed (check!)

$$\widehat{\phi}(\omega) = \frac{1 - e^{-3i\omega}}{3i\omega}$$

and this transform has a 2π-periodic zero (where?) □

Construction of Daubechies' refinable functions.

Let k be a positive integer. Consider the binomial expansion of

$$(\cos^2(\omega/2) + \sin^2(\omega/2))^{2k-1}, \tag{13}$$

and order the terms in decreasing powers of cos (i.e., the first term is $\cos^{4k-2}(\omega/2)$). Let

$$T(\omega)$$

be the sum of the first k terms in this expansion. For example, when $k = 2$,

$$T(\omega) = \cos^6(\omega/2) + 3\cos^4(\omega/2)\sin^2(\omega/2). \tag{14}$$

$T(\omega)$ is a trigonometric polynomial (why?). It is also clear that $T(\omega) \geq 0$ for every ω, and that $T(0) = 1$. Finally, we observe that $T(\omega + \pi)$ is the sum of the *last* k summands in (13), and hence

$$T(\omega) + T(\omega + \pi) = 1.$$

The Féjér-Riesz Lemma 15. *Let* T *be a trigonometric polynomial which is non-negative everywhere. Then there exists a trigonometric polynomial* H *such that*

$$T(\omega) = |H(\omega)|^2.$$

Applying this lemma, we obtain a trigonometric polynomial, H_k, such that $|H_k(\omega)|^2 = T(\omega)$. For example, for the case (14) this polynomial turns out to be

$$H_2(\omega) = \cos^2(\omega/2)(\frac{1 + \sqrt(3)}{2} + \frac{1 - \sqrt(3)}{2}e^{i\omega}).$$

Note that we can now conclude that each of the above H_k functions is a CQF. Then, Theorem 12 implies most of next result.

Theorem and Definition. *For each of the above trigonometric polynomials H_k, there exists a function D_k (known as 'Daubechies' refinable function' of order k) such that*

(i) *D_k is refinable with mask H_k.*
(ii) *D_k is supported in the interval* $[0, 2k-1]$.
(iii) *The shifts $E(D_k)$ of D_k are orthonormal.*
(iv) *The mask h_k associated with H_k has exactly $2k$ non-zero coefficients:*

$$H_k(\omega) = \sum_{m=0}^{2k-1} h_k(m) e^{-im\omega}.$$

Note that there are two pieces missing in the above theorem. The first concerns the actual 'computation' of the function D_k. It turns out that: (a) this can be done with ease (using the tool of the *cascade algorithm* that will be discussed in the sequel). (b) It is not important at all: the entire practical implementation of wavelets can be done purely in terms of *masks*. This will become clear as soon as we discuss the construction of wavelets, and the fast wavelet transform.

The other issue is the smoothness of D_k, an issue of critical importance. Estimating the smoothness of refinable functions (by inspecting their masks) is a formidable problem, and is among the hardest problems in the theory of wavelets. The major success of Daubechies' construction was her ability to prove the following celebrated result:

Theorem 16. *For each positive integer k, one can find a positive integer k' such that the refinable function $D_{k'}$ has k continuous derivatives.*

The exact connection between k and k' is rather complicated. For large values of k, we have approximately that $k' \approx 5k$. On a more practical level, D_2 (which is supported in an interval on length 5), can be proved to have (barely) one continuous derivative.

Part 6: MRA and the unitary extension principle

We now discuss how wavelet systems are derived from refinable functions.

We start by selecting a refinable function ϕ, and we denote by H_0 its refinement mask:

$$\widehat{\phi}(2\omega) = H_0(\omega)\widehat{\phi}(\omega).$$

Next we denote by

$$V_1 := V_1(\phi)$$

the space 'spanned' by $(\phi_{1,k})_{k\in\mathbb{Z}}$ (more precisely: V_1 is the L_2-closure of the finite span of $(\phi_{1,k})_{k\in\mathbb{Z}}$.)

Examples. Let ϕ be B_1, the B-spline of order 1. Then $\phi_{1,k}$ is (up to multiplication by a normalization constant) the support function of the interval $[k/2, (k+1)/2]$. V_1 is then the space of all functions in L_2 which are piecewise-constants with (possible) breakpoints at the half-integers.

Similarly, if we choose ϕ to be the hat function B_2, then V_1 becomes the space of all functions in L_2 which are continuous piecewise-linear with (possible) breakpoints at the half integers. □

Note that the refinability assumption on ϕ implies that $\phi \in V_1$. Note that the space V_1 is invariant under shifts, i.e., if $f \in V_1$ then $E^k f \in V_1$ for every *integer* k (in fact, that is true even for $k \in \mathbb{Z}/2$, but we do not need it here.) Thus, we actually have that

$$E(\phi) \subset V_1.$$

Our goal is to construct a wavelet system. Recall that constructing a wavelet system in tantamount to selecting the *mother wavelets* Ψ. We will select all the mother wavelets from the space V_1. In some sense, the mother wavelets $\Psi \subset V_1$ 'complement' ϕ.

Discussion. One might equate the situation here to a simple setup in Linear Algebra: we are given vector v, and attempt to extend this vector to a spanning set of a given vector space. Here, the role of the vector v is played by the shifts $E(\phi)$ of the refinable function ϕ, and the underlying vector space is V_1. We complement $E(\phi)$ with the selection of $E(\Psi) \subset V_1$. □

Let ψ be *any* function in V_1. Then one can prove that there exists a 2π-periodic function H_ψ such that

$$\widehat{\psi}(\omega) = H_\psi(\omega/2)\widehat{\phi}(\omega/2). \tag{17}$$

This should come at no surprise: if we write

$$H_\psi(\omega) := \sum_{k=-\infty}^{\infty} h_\psi(k) e^{-ik\omega},$$

then a simple exercise yields that (17) implies that

$$\frac{1}{2}\psi(t) = \sum_{k\in\mathbb{Z}} h_\psi(k)\phi(2t-k). \tag{18}$$

Since the function $t \mapsto \phi(2t-k)$ is (up to normalization by $\sqrt{2}$) the function $\phi_{1,k}$, (18) says that $\psi \in V_1$. (One should be warned that (18) may not always make sense, while the fact that every function in V_1 satisfies an equation of the form (17) is always true).

Interim Summary. Our objective is to select mother wavelets from the space $V_1(\phi)$ associated with a refinable function ϕ. The refinable function is completely determined by its refinement mask, which we denote here by H_0. Selecting mother wavelets $(\psi_m)_{m=1}^N \in V_1$ is on par with selecting 2π-periodic functions $(H_m)_{m=1}^N$ (and then defining the wavelet ψ_m by the relation (cf. (17))

$$\widehat{\psi_m}(\omega) = H_m(\omega/2)\widehat{\phi}(\omega/2). \tag{19}$$

So, the MRA construction of wavelet systems is reduced to the following setup:

The MRA setup: Let H_0 be a given 2π-periodic (bounded) function, satisfying $H(0) = 1$. Select 2π-periodic bounded functions $H_1, \ldots, H_N$. Then: with ϕ the refinable function associated with H_0, you obtain a wavelet system whose mother wavelets $\psi_1, \ldots, \psi_N$ are defined by (19). We refer below to such a construction as **the wavelet system associated with the refinement mask H_0 and the wavelet masks $(H_1, \ldots, H_m)$.** □

The ultimate goal is to understand the process well-enough so that we can construct in this way *good* wavelet systems. The immediate question is thus what we mean by 'good'. The five basic desired properties are as follows:

(i) Localness in time: ideally we would like all the mother wavelets to be supported in a small interval.

(ii) Smoothness: in order to be local in frequency, all the wavelets should be as smooth as possible.

(iii) High *vanishing moments*: we would like each $\widehat{\psi}_m$ to have a high-order zero at the origin (this is the other condition for frequency localization).

(iv) Good reconstruction method. Ideally, we would like the wavelet system to satisfy the *perfect* reconstruction formula:

$$f = \sum_{\psi,k,j} \langle f, \psi_{j,k} \rangle \psi_{j,k}, \qquad \forall f \in L_2.$$

(v) symmetry, or anti-symmetry. We postpone a discussion of that issue.

We ignore, for the time being, properties (iii) and (v). Note that property (ii) is completely controlled by the choice of ϕ (i.e., the choice of H_0): once ϕ is smooth, all the functions in V_1 will be smooth, too. Property (i) is also largely controlled by the choice of ϕ: if ϕ is compactly supported, we will get compactly supported wavelets by simply choosing all the functions (H_m) to be trigonometric polynomials (show that!).

The crux in the MRA construction of wavelets is property (iv). Let's name that property first:

Definition 20. *Let G be a system of functions in L_2. We say that G is a* **tight frame** *for L_2 if the perfect reconstruction property is valid:*

$$f = \sum_{g \in G} \langle f, g \rangle \, g. \qquad \forall f \in L_2.$$

□

We study this property further in a later section. At this point, one may interpret a tight frame as some relaxation of the notion of orthonormal system.

Theorem 21 (The unitary extension principle). *Let ϕ be refinable with mask H_0, $H_0(0) = 1$. Let $H_1, \ldots, H_N$ be 2π-periodic functions. Assume that the following two conditions hold for every $\omega \in [-\pi, \pi]$:*

$$|H_0(\omega)|^2 + |H_1(\omega)|^2 + \ldots + |H_N(\omega)|^2 = 1, \tag{22}$$

and

$$H_0(\omega)\overline{H_0(\omega + \pi)} + H_1(\omega)\overline{H_1(\omega + \pi)} + \ldots + H_N(\omega)\overline{H_N(\omega + \pi)} = 0. \tag{23}$$

Then the wavelet system associated with the refinement mask H_0 and the wavelet masks $(H_1, \ldots, H_N)$ is a tight frame.

Example. We examine the binomial expansion of

$$(\cos^2(\omega/2) + \sin^2(\omega/2))^2, \tag{24}$$

and take H_m to be the square root of the mth term in that expansion. Thus,

$$H_0(\omega) := \cos^2(\omega/2), \quad H_1(\omega) := i\sqrt{2}\cos(\omega/2)\sin(\omega/2), \quad H_2(\omega) := \sin^2(\omega/2).$$

Then condition (22) is obviously valid here, and condition (23) can be easily verified (the three terms that are obtained when checking that latter condition come from the expansion of

$$(\cos(\omega/2)\sin(\omega/2) - \cos(\omega/2)\sin(\omega/2))^2.)$$

So, the above should lead to a tight wavelet frame. Let's find the two mother wavelets. First, H_0 is known to be the mask of the (centered) hat function. In order to find the wavelets, we first find the Fourier coefficients h_1, h_2 of H_1 and H_2. Note

$$H_1(\omega) = i\frac{\sqrt{2}}{2}\sin(\omega) = \frac{\sqrt{2}}{4}(e^{i\omega} - e^{-i\omega}).$$

This means that $h_1(1) = \frac{-\sqrt{2}}{4}$, $h_1(-1) = \frac{\sqrt{2}}{4}$, and

$$\psi_1(t) = 2(h_1(1)\phi(2t-1) + h_1(-1)\phi(2t+1)) = \frac{\sqrt{2}}{2}(\phi(2t+1) - \phi(2t-1)).$$

Similarly,

$$H_2(\omega) = \sin^2(\omega/2) = \frac{1-\cos(\omega)}{2} = \frac{-e^{-i\omega} + 2 - e^{i\omega}}{4}.$$

This time $h_2(-1) = h_2(1) = -\frac{1}{4}$, while $h(0) = \frac{1}{2}$, hence

$$\psi_2(t) = 2(h_1(1)\phi(2t-1) + h_0(0)\phi(2t) + h_1(1)\phi(2t+1)) = 1 - \frac{1}{2}(\phi(2t-1) + \phi(2t+1)).$$

The above example generalizes to higher order B-spline. One just needs to use a higher power in (24). Note that the number of wavelet increases together with that power. □

There is no general recipe for constructing a tight wavelet frame based on the unitary extension principle. Such a rule does exist, however, if the refinement mask is a CQF.

Construction of a wavelet system from a CQF mask: Mallat's construction Since a CQF mask H_0 satisfies

$$|H_0(\omega)|^2 + |H_0(\omega + \pi)|^2 = 1,$$

one may be tempted in this case to choose a single wavelet mask $H_1(\omega) := H_0(\omega + \pi)$, since then (22) is trivially satisfied. However, this is too crude for the satisfaction of (23). Instead, we define the unique wavelet mask to be

$$H_1(\omega) := e^{i\omega}\overline{H_0(\omega + \pi)}.$$

Then (22) and (23) are always satisfied, and we obtain a tight wavelet frame generated by a single mother wavelet. The name CQF is usually connected to the *pair* (H_0, H_1).

Let's try to decipher the meaning of

$$H_1(\omega) := e^{i\omega}\overline{H_0(\omega + \pi)}.$$

Suppose that we know the sequence (filter) h_0 whose Fourier series is H_0. How do we modify the filter h_0 in order to obtain the filter h_1? There are three easy steps:

(1) Shifting H_0 by π: this amounts to changing the signs of all the coefficients at odd locations, i.e., we replace $h_0(k)$ by $(-1)^k h_0(k)$.

(2) Applying complex conjugation to $H_0(\omega + \pi)$: this amounts to interchanging the positive location with the negative location, so that we now find at the kth location $(-1)^{-k} h_0(-k)$.

(3) Multiplying the result by $e^{i\omega}$: this amounts to shifting the filter one step (forward or backward, as you wish: we could have defined H_1 with $e^{-i\omega}$ instead of $e^{i\omega}$). Thus, finally,

$$h_1(k) = (-1)^{k-1} h_0(-k + 1).$$

Example. The mask H_0 of the refinable B_1 is associated with the filter $h_0(0) = h_0(1) = 1/2$. Its mirror filter h_1 is thus defined by $h_1(0) = -\frac{1}{2}$, $h_1(1) = \frac{1}{2}$. Note that the resulting wavelet is (no surprise) the Haar wavelet.

We conclude this part with the following improvement of the CQF construction:

Theorem 25. *Let (H_0, H_1) be a CQF pair. If the shifts $E(\phi)$ of the underlying refinable function are orthonormal then the corresponding wavelet system is complete and orthonormal.*

Example. The shifts of B_1 are indeed orthonormal, and the resulting Haar wavelet system is indeed complete and orthonormal. The same applies to each of Daubechies' refinable functions: the resulting wavelet system is complete and orthonormal. In contrast, while the mask of the support function ϕ of the interval $[0, 3]$ is a CQF, $E(\phi)$ is not an orthonormal system. Indeed, the wavelet system constructed by applying the CQF unitary extension principle is a tight frame (as it should), but is not an orthonormal system. □

Part 7: Filters, filter banks, and the fast wavelet transform

We would like first to connect the theoretical discussion so far, to practical algorithms. We start with the notions of *signals*, *filters*, and *low-pass/high-pass* filters.

Discussion: filters. In signal analysis, functions cannot be given as a continuum of values. Instead, we are given a discrete sequence of values which we can index by the integers

$$k \mapsto x(k), \quad k \in \mathbb{Z},$$

and refer to as a **signal**. The signal x can be either obtained by sampling some given function, or by some other (local) processing of the function. We need to assume that the process is regular in time, e.g., in the case of sampling this means that we have sampled the underlying function equidistantly in time.

With only discrete information on f in hand, we need to discretize some of the operations we use. In a natural way, the Fourier transform is replaced by the Fourier series:

$$X(\omega) := \sum_{k=-\infty}^{\infty} x(k)e^{-ik\omega}.$$

Next, *convolution* is replaced by discrete convolution:

$$(h * x)(k) := \sum_{m\in\mathbb{Z}} h(m)x(k-m).$$

Note that the Fourier series of $h * x$ is the function $H(\omega)X(\omega)$. While the action of convolution is commutative ($h * x = x * h$), the user will usually regard differently the two sequences: one, say x, is the given signal. The other, h, is an especially designed sequence, made in order to separate (i.e., 'filter', 'mask') certain properties of x; first, and foremost, frequency properties. For that reasons, h (or more precisely the convolution action $h* : x \mapsto h * x$ (which acts on all signals) is referred to as **filter**. A filter h is **low-pass** if H is concentrated around the origin (and therefore vanishes at the 'end points' $\pm\pi$). Note that this means that $H_1(\omega) := H(\omega + \pi)$ vanishes at the origin, and is concentrated at the end points $\pm\pi$. Its corresponding filter h_1 is thus a **high-pass filter**. Computing the sequence h_1 in terms of h is easy, and one finds that

$$h_1(k) = (-1)^k h(k).$$

So we have just found an easy way to associate low-pass filter with high-pass filter and vice versa.

Filtering a given signal x (i.e., replacing it by $h * x$) results in the enhancement of certain properties of x, and in the suppression of others. It is rather hard to recover the original signal x from its filtered version $h * x$. For example, if h is a very good low pass filter, then H is very flat at the origin, very flat at π, and $H(0) = 1$ while $H(\pi) = 0$ (and what about $-\pi$?). This means that the filtering by h results in a signal whose Fourier series $H(\omega)X(\omega)$ preserves very accurately the low frequency content of x while suppresses completely the frequencies of x near π. It will be very hard, thus, and highly non-robust to recover x from $h * x$.

In order to address this problem, one can use several, complementary filters. Say, a low-pass h_0 and a high pass h_1. The immediate problem is then of *oversampling*: if the filter h is short (i.e., has only a few non-zero values), then the size (i.e., the number of non-zero values) in $h * x$ is on par with those of x. However, if we use 2 filters h_0, h_1, we find ourselves dealing with a combined 'processed' signal of size double the original one.

Heuristically, one then should guess that some of the values in $h_0 * x, h_1 * x$ should be discarded. For example, why not discard every other sample. This leads to the operation of **downsampling**:

$$x_{\downarrow}(k) := x(2k).$$

Note that $x_{\downarrow}$ preserves only the values of x at even locations (and renumber those locations).

Decomposition of a signal using filter banks. We restrict our attention to the setup that is connected to decomposition by wavelet systems. Let $(h_0, h_1, \ldots, h_N)$ be a filter bank. We assume that h_0 is a low-pass filter, and all the others are high-pass filters. Set:

$$\nu_{1,m} := \sqrt{2}(h_m * x)_{\downarrow}.$$

Note that if we have more than one high-pass filter, we still oversample (by how much?). This is the *intrinsic* oversampling of the process.

We then proceed by reapplying the process to $\nu_{1,0}$ (which corresponds to the low-pass filtering of x). We do not touch any more $\nu_{1,m}$, $m > 0$. (There are applications where it is necessary to reprocess the high-frequency components of x. The wavelet theory that relates to these algorithms is connected with the notion of *wavelet packets*. We will not discuss it here).

Thus, in the next stage we decompose $\nu_{1,0}$. Inductively, we define:

$$\nu_{j,m} := \sqrt{2}(h_m * \nu_{j-1,0})_{\downarrow}. \tag{26}$$

Note that our labeling of the frequency grades is opposite to that used in the wavelet: $\nu_{2,m}$ corresponds to frequencies *lower* than $\nu_{1,m}$.

The connection between filter banks, the above process and wavelet system is given in the next (easy to prove) theorem:

Theorem: the fast wavelet/frame transform. *Let f be some function, let ϕ be some refinable function, and denote:*

$$x(k) := \langle f, E^k \phi \rangle, \qquad k \in \mathbb{Z}.$$

Let H_0 be the refinement mask of ϕ, and let W_Ψ be the wavelet system associated with the refinement mask H_0 and the wavelet masks $(H_1, \ldots, H_N)$. Let $h_0, \ldots, h_N$ be the corresponding filters. Then, in the notation of (26), and for every $j > 0$,

$$\langle f, \psi^m_{-j,k} \rangle = \nu_{j,m}(k), \quad m = 1, \ldots, N,$$

with ψ^m the wavelet associated with the mask H_m. Moreover,

$$\langle f, \phi_{-j,k} \rangle = \nu_{j,0}(k).$$

Part 8: Reconstruction.

The reconstruction goal can be described as follows:

Given a collection of linear functionals Λ, associate each one of them with a function g_λ such that we obtain the perfect reconstruction formula:

$$f = \sum_{\lambda \in \Lambda} \langle f, \lambda \rangle g_\lambda, \quad \forall f.$$

In the case Λ is orthonormal, and even in the case Λ is a tight frame, we can take $g_\lambda = \lambda$. There are many interesting aspects to the more general case, when we reconstruct using a system different from the one we used to decompose. This will be discussed in the next section. Our direction in this section is rather different.

We want, in view of the development of filter banks and the fast wavelet transform, and in view of the fact that our actual world is discrete, to re-examine the notion of reconstruction.

We have seen in the last section that 'decomposition' in the practical level, does not mean that we actually compute the inner products of a given function against the elements Λ of the system. Rather, we assume that we are given the inner products of f with respect to some system, and decompose those inner products. It is rather ambitious thus to attempt at finding the actual function f during the reconstruction, while we did not assume to have full access to the function in the first place.

Instead, we merely should wish to invert the process of decomposition. The fast wavelet transform produces sequences of the form

$$\nu_{j,m}, \quad j = 1, 2, \ldots, J \quad m = 1, \ldots, N.$$

Note that we are assuming that we terminated the decomposition process after J steps. This means that we need to retain also the lowest frequency part of the signal

$$\nu_{J,0}$$

since this part was not decomposed further (note that at previous levels we retain only the high-frequency values, which correspond indeed to the inner products with the wavelets).

Our theorem concerning the unitary extension principle leads to a tight wavelet frame. The tight frame property says that we should be able to reconstruct using the same wavelet system that we used to decompose. In terms of the masks and its filters, this should indicate that we might be able to use (essentially) the same masks during the reconstruction.

The reconstruction algorithm is recursive:

`for j=J:-1:0`
`use` $\nu_{j,m}$`,` $m = 0, 1, \ldots, N$`,`
`in order to reassemble the sequence` $\nu_{j-1,0}$
`end`

In order to understand the reconstruction process, it is instructive to envision the decomposition part of the fast wavelet transform on the frequency domain. On the time domain, we decomposed $\nu_{j-1,0}$ as follows:

$$\nu_{j,m} = \sqrt{2}(h_m * \nu_{j-1,0})_\downarrow, m = 0, \ldots, N. \tag{27}$$

Let's denote by

$$X_{j,m}$$

the Fourier series of $\nu_{j,m}$. With some (but not much) effort, one shows that (27) can be rewritten on the frequency domain as

$$X_{j,m}(\omega) = (H_m X_{j-1,0})(\omega/2) + (H_m X_{j-1,0})(\omega/2 + \pi).$$

Now, substitute 2ω for ω, and then multiply each side of the last equality by $\overline{H_m}(\omega)$, and sum over all m. Then

$$\sum_{m=0}^{N} \overline{H_m(\omega)} X_{j,m}(2\omega) = (\sum_{m=0}^{N} |H_m(\omega)|^2) X_{j-1,0}(\omega) + (\sum_{m=0}^{N} (\overline{H_m} H_m)(\omega+\pi) X_{j-1,0}(\omega+\pi).$$

Now, comes the punch-line: if our filter bank satisfies the unitary extension principle, we can use (22) and (23) to conclude that

$$\sum_{m=0}^{N} \overline{H_m(\omega)} X_{j,m}(2\omega) = X_{j-1,0}(\omega).$$

This means that we found a reconstruction algorithm; we only need (if we want to implement the algorithm on the time domain) to understand, on the time domain) the meaning of

$$\overline{H_m(\omega)} X_{j,m}(2\omega).$$

There are two actions here:

(1) Dilation: $X_{j,m}(\omega) \mapsto X_{j,m}(2\omega)$. This is simply a relabelling of the entries of the signal $x_{j,m}$. If we define the **upsampling operator**

$$x_{\uparrow}(k) = \frac{1}{\sqrt{2}} \begin{cases} x(k/2), & k \text{ is even}, \\ 0, & \text{otherwise}, \end{cases},$$

then, with $y := x_{\uparrow}$, $Y(\omega) = X(2\omega)$. The switch from $H_m(\omega)$ to $H_m(\omega/2)$ is just a relabeling of the coefficients in *to be continued...*

Part 10: More on refinable functions

When constructing wavelets via MRA, the choice of the refinable function plays a major role:

(1) It determines completely the smoothness of the mother wavelets (why?)

(2) It determines almost completely the localness in time of the wavelet: it is practically impossible to construct a wavelet system with compactly supported mother wavelets unless the corresponding refinable function is compactly supported.

(3) It determines to a large degree the number of vanishing moments the mother wavelets have. Remember that we say that ψ has m **vanishing moments** if

$$\widehat{\psi}^{\ell}(0) = 0, \quad \ell = 0, \ldots, k-1.$$

Recall that the frequency localization of the wavelet is determined its smoothness and its vanishing moments.

(4) It determines to some degree the properties of the resulting wavelet system. For example, if the shifts of the refinable function are orthonormal, we can construct (via MRA) an orthonormal wavelet system (using e.g., the unitary extension principle).

In summary, we need to be able to construct refinable functions with desired properties. We list some of these desired properties:

(1) Smoothness: we would like to have a smooth refinable function.

(2) High approximation order: we explain that property later. This is the property of the refinable function that we allow us to generate wavelets with high vanishing moments.

(3) Orthonormality (or a similar property) of the shifts $E(\phi)$.

We must, therefore, keep in mind that the refinable function, almost always, is not given to us in an explicit form. We choose the refinement mask H_0, and need to know how to read the desired properties of ϕ from its refinement mask H_0

Before we turn our attention to this problem, we ask a simpler one: given the mask H_0, is there a simple way to visualize the corresponding refinable function ϕ? An affirmative answer is given in terms of

The Cascade Algorithm. A refinable ϕ with mask h_0 satisfies (by definition) the refinement relation

$$\phi(t) = 2 \sum_{k \in \mathbb{Z}} h_0(k) \phi(2t - k).$$

Now, let us define the **Cascade operator**

$$C(f)(t) := \sum_{k \in \mathbb{Z}} h_0(k) f(2t - k).$$

Thus the cascade operator maps a give function f to a linear combination of the dilated shifts of that function. We have chosen the coefficients in that linear combination to be those of the refinement equation. Thus,

$$C(\phi) = \phi.$$

In the language of linear algebra, ϕ is an *eigenvector* of C. In the language of Numerical Analysis, ϕ is a *fixed point* of C. A standard way to attempt finding a fixed point is by *iterations*:

`Starting with some initial function` ϕ^0`,`

`define` $\phi^m := C(\phi^{m-1})$`,` $m = 1, 2, \ldots$.

It turns out that the algorithms succeeds only if the initial function ϕ^0 partition unity in the sense that

$$\sum_{k \in \mathbb{Z}} \phi^0(t - k) = 1.$$

Such functions exist in abundance. For example, all the B-splines satisfy this property. A standard choice for ϕ^0 is the centered hat function.

Does the cascade algorithm converge? What does it mean 'to converge' here? Can it converge to function other than the refinable ϕ?

Fact 28. *If the mask H_0 is a trigonometric polynomial, and if $H_0(0) = 1$, then the cascade algorithm either diverges, or converges to the refinable ϕ. It does not matter in that context how exactly 'convergence' is defined.* □

Definition of 'convergence'. There are several possible definition here. One of those is as follows:

$$\|\phi - \phi^m\|_{L_2} \to 0,$$

as $m \to \infty$.

While the complete characterization of the convergence of the cascade algorithm is non-trivial, there are important cases where such convergence is guaranteed:

Theorem 29. *If the shifts of the refinable ϕ are orthonormal, or even if they only form a Riesz basis (a notion that is defined in the next section), the cascade algorithm converges.*

One should be warned that the cascade algorithm may fail to converge if we only know that the mask H_0 of the refinable ϕ is a CQF. An example of that possible phenomenon is given the support function of the interval $[0, 3]$.

How to determine properties of the refinable function from its mask H_0?

We focus on three basic properties: the smoothness of ϕ, the approximation order the shifts of ϕ, and the (possible) orthonormality of the shift of ϕ.

Part 10: Good Systems

(6:) Summary: a general recipe for constructing the linear functionals of the analysis map Λ^*.

Step I: select a suitable 'window' function g (more generally, select a few such windows). The window function is always selected with great care (and much of the theory goes into the question of how to construct useful window functions). The window function induces the linear functional

$$\lambda_g : f \mapsto \langle f, g \rangle := \int_{\mathbb{R}} f(t)\overline{g(t)}\, dt.$$

Notions in the context of the current discussion:

(a) **Compactly supported functions**: a function $f : \mathbb{R} \to \mathbb{R}$ is compactly supported if it is identically 0 outside of some bounded closed interval (the smallest such interval is the **support** of f; warning: there are finer definitions for the notion of support).

Examples of compactly supported functions: the function B_1, B_2, B_3, $H_{j,k}$ $(0 \le j, k \le 1)$ from Assignment 1.

(b) **Periodic functions:** A function $f : \mathbb{R} \to \mathbb{R}$ is periodic (with period 2π) if

$$f(t + 2\pi) = f(t), \quad \forall t \in \mathbb{R}.$$

Examples of such functions are abundant. E.g., for every integer (positive or negative, or even 0) number, each of the functions

$$t \mapsto sin(nt), \quad t \mapsto cos(nt)$$

is 2π-periodic. Closely related to those are the periodic exponential functions

$$e_n : t \mapsto e^{int} = cos(nt) + i \sin(nt).$$

The *ideal window* is a periodic exponential function supported in an interval of length zero. Obviously, such function does not exist (it is not only that there exists no function supported at a single point. In fact we could interpret the point evaluation functional as 'a function supported at one point'; however, the periodic exponential functions are far from having one point support, and none of them is even compactly supported. moreover, the only compactly supported periodic function is the 0-function). For reasons that will be explained later, we try to 'get close' to the ideal window by constructing window functions that are *local* (i.e., supported in a small interval) and *smooth* (i.e., possess many continuous derivatives; look at the above examples of compactly supported functions, to realize that compactly supported functions may have very low smoothness). In this regard, it is useful to recall the notation (for a non-negative integer k)

$$C^k(\mathbb{R}) := \{f : \mathbb{R} \to \mathbb{R} : \text{ the } k\text{th order derivative of } f \text{ exists and is continuous}\}.$$

(The case $k = 0$ refers to *continuous functions*. If $f \in C^k(\mathbb{R})$ for every k, we write $f \in C^\infty(\mathbb{R})$, and we say that f is *infinitely differentiable*. Note that we never differentiate any function infinitely many times, despite of the above name).

We also recall that in the context of Fourier analysis we measure the smoothness not in terms of *continuous* derivatives, but in terms of derivatives that lie in L_2. Thus we have another space

$$W^k(\mathbb{R}) := \{f : \mathbb{R} \to \mathbb{R} : \text{ all the derivative of } f \text{ up to order } k \text{ exist and are in } L_2\}.$$

Example. Let B_2 be the hat function. The hat function is continuous, but its derivative is not. Therefore, $B_2 \in C^0$, but $B_2 \notin C^1$. On the other hand, the first derivative of B_2 is still in L_2, hence $B_2 \in W^1$ (but not in W^2). □

Step II:

After you selected you window function(s), you select the operation(s) you would like to apply to this window:

(1) translation

(2) modulation

(3) dilation

The different choices have beautiful names:

translation	$\Longrightarrow$	convolution
modulation	$\Longrightarrow$	fourier transform (here g is the constant function)
translation+modulation	$\Longrightarrow$	Gabor system
translation+dilation	$\Longrightarrow$	wavelet system

(7:) so what does it mean to be a 'good system'?

A 'good system' does not relate necessarily to the operations used to produce the system. Wavelet systems are neither better nor worse than Gabor systems. They simply fit different applications and have different theories and different algorithms. The notion of a 'good system' is universal to all the systems.

There are two basic criteria, which are seemingly unrelated (but are, as a matter of fact very much related) that guide us in classifying 'good systems'.

(I) We want to have a close relation between the 'size' of the function f we analyse, and the 'size' of the numbers we produce via $f \mapsto \Lambda^* f$.

Parseval's identity tells us that, in the context of Fourier analysis we do achieve such a relation. In fact, this is the case for every complete orthonormal system.

We measure the size of f by its L_2-norm

$$\|f\| := \|f\|_{L_2(\mathbb{R})}.$$

and we measure the size of $\Lambda^* f$ by the ℓ_2-norm:

$$\|\Lambda^* f\| := (\sum_{i=0}^{\infty} |\lambda_i(f)|^2)^{1/2}.$$

We want then the norm of $\Lambda^* f$ to be 'nicely' related to the $\|f\|$. Here are the relevant definitions: