Amos Ron
Lectures Notes, CS515
4Mar05

CS515, S05, lecture notes

Amos Ron

Note on notes: *these notes are meant to be readable only by the lecturer. They are being put in public domain since it is possible that parts of them might be helpful to others (e.g., students at CS515).*

0. Welcome to CS515

Data, data, data !!!! The last decade was marked by breathtaking increases in data communications and acquisition abilities, and in data processing power. In today's world, questions concerning data, their representation, manipulation, and analysis are proved critical to technological advance: from high energy physics to astronomy; from analysis of financial data to deeper understanding of the Internet dynamics; from wireless communication technologies to computer graphics applications; from microscopy technologies to biological and medical studies of the human body, the brain and the genetic system. Understanding the art of data processing is one of the most valuable commodities in today's life (read: taking cs515 is one of the smartest decisions that you have ever made!)

The first half of this course is devoted to developing the basic theory of *time-frequency* techniques of *data representation*. This theory is referred to colloquially as "wavelet theory" and was developed mostly in the last two decades. In order to begin, one needs to understand first the notion of "data representation", which, by and large, is the most basic principle in any data processing approach. Then, in order to understand "time-frequency representation" one has to be acquainted first, separately, with the "time representation" and the "frequency representation" of data. And, how does mathematics realize "data" to begin with? Addressing the issues raised this paragraph is the prelude to wavelet theory, and will be the subject of the first few lectures. Let us then begin the journey.

Before we formalize the notion of "data representation" it is instructive to look at some examples. On my laptop there are a few:

(1) a musical signal, represented in four different forms (as notes, as a time signal, as a frequency signal, and one of its possible time-frequency representations)
(2) an image is decomposed into 36 small images
(3) an Internet traffic signal and its framelet decomposition

Mathematics treats "data" as functions. The simplest of which are functions in one dimension:

$$f : \mathbb{R} \to \mathbb{R} : t \mapsto f(t).$$

Note that we assume the function f to be defined on the entire real line and to assume real values. The latter is convenient, but from time to time we will also consider functions that assume complex values, i.e., whose target is $\mathbb{C}$. Many functions of interest are not defined on the entire real line, but only on a closed interval $[a, b]$. For our mere convenience, we will extend those functions to the entire line by defining them to be zero outside their interval of definition. We refer to the independent argument, t, of the function f as *time*, regardless whether this makes practical sense or not.

However, the functions that we are interested in this course are very different from those that a student may encounter, say, in a calculus course. They are usually very oscillatory, with hundreds and perhaps thousands of local extrema. We usually look for hidden "features" in these functions, but in many instances we do not know a priori how to define correctly the notion of a "feature", neither in mathematical terms nor in any other, less abstract, ones.

While the functions that we are mainly interested in are *real-valued* (i.e., their target is the real line $\mathbb{R}$), our theory and analysis employs complex numbers *everywhere*. It is useful to keep in mind some of complex numbers basics. A few basics are listed here (we use i for the imaginary $\sqrt{-1}$). Here, $\omega, a, b \in \mathbb{R}$, while $x, y \in \mathbb{C}$.

$$|a + bi| = \sqrt{a^2 + b^2}, \quad |xy| = |x||y|, \quad e^{i\omega} = cos(\omega) + i\sin(\omega), \quad |e^{i\omega}| = 1,$$

$$\overline{a + bi} = a - bi, \quad x\overline{x} = |x|^2, \quad \overline{e^{i\omega}} = e^{-i\omega}.$$

Part 1. Introduction: Representation

At the heart of data representation (at least at the heart of the part of the theory that we address in this course) is the notion of a *linear functional*. Let us assume first that our objective is not the study of a single function f, but rather a family of functions F (each of which is still defined on the real line etc.) Let us assume further that F is not just a "collection of functions" but a *space of functions*, in the sense of Linear Algebra (i.e., a vector space), i.e.,

$$f, g \in F,\ a \in \mathbb{R} \Longrightarrow f + g \in F,\ af \in F.$$

Once we fix F, we say that a map

$$\lambda : F \to \mathbb{R}$$

is a **linear functional** if it is a linear map, i.e., if the following two conditions are satisfied:

(a) For every $f, g \in F$, $\lambda(f + g) = \lambda(f) + \lambda(g)$, and
(b) For every $f \in F$ and $a \in \mathbb{R}$, $\lambda(af) = a\lambda(f)$.

Intuitively, the linear functional is a "feature", or a "feature detector"; if, for given function f, $\lambda(f) = 0$ one might think about that as " the feature represented by λ does not appear in the function f". Let us try to make this more concrete by examining specific functionals.

Examples.

(1) Choose some (fixed) $t \in \mathbb{R}$ and define $\lambda_t(f) := f(t)$. This is the linear functional of *point evaluation*. The feature captured by λ_t is the most obvious one: the value of the function at t.

(2) Choose some (fixed) $t \in \mathbb{R}$ and define $\lambda(f) := f'(t)$ (for this λ we need all the functions in F to be differentiable at t.) This is the linear functional of *derivative evaluation*.

(3) Fix any function $g : \mathbb{R} \to \mathbb{R}$ (which might be in F or might not be there). Define:

$$\lambda_g(f) := \int_{\mathbb{R}} f(t)\overline{g(t)}\, dt (= \int_{-\infty}^{\infty} f(t)g(t)\, dt).$$

In our discussion, we will actually consider linear functionals of this class only. It is now far less trivial to identify the "feature captured by the linear functional λ_g". Sometimes a good way to realize this feature is to look at the graph of g. Sometimes, we invoke more sophisticated techniques. A part of the art here is to design g so that λ_g "captures the feature we are interested in". We will go into great length later in expanding and providing a concrete context to the above general, abstract, discussion.

Examples for (3) above include local averaging and local differencing. Given some fixed interval $[a, b]$, the *local averaging* g over that interval is the function

$$g(t) := \begin{cases} \frac{1}{b-a}, & a \le t \le b, \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

A local differencing is obtained by subtracting two local averaging functions.

The notion of local averaging can be generalized, and the generalization is actually very useful. For that, we first would like to recall the notion of *compact support*.

Definition: a function with compact support. Let $g : \mathbb{R} \to \mathbb{R}$ be given. We say that g **has compact support** (or is "compactly supported") if g vanishes everywhere outside of some bounded interval $[a, b]$. The smallest such interval is called the **interval support** of the function g. □

The local averaging in (1) is compactly supported. Its interval support is the interval $[a, b]$ that appears there. In addition to the compact support, the local averaging (1) has the additional property of having *mean value 1*, i.e.,

$$\int_{\mathbb{R}} g(t)\, dt = 1.$$

Definition. Every function with compact support and mean value 1 is called local averaging. □

The notion is, needless to say, very qualitative. There is tremendous difference between an averaging function whose interval support is $[0, 1]$, and another averaging function whose interval support is $[0, 1000]$.

In the same spirit of the general local averaging, one can talk about general local differencing: if g and h are two averaging functions, their difference

$$g - h$$

is a (generalized) local differencing. We sometimes refer to a local differencing function as *discrete differentiation.*

(2) The analysis map. We use linear functionals as the soldiers in the assembly of a real army. This real army if the analysis map (also known as the decomposition map), which provides the given function f with a new representation.

To this end, we let Λ be a family of linear functionals indexed by some index set I (which may be finite or infinite):

$$\Lambda := \{\lambda_i\}_{i \in I}.$$

The map

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

in the analysis map. We often use the positive integers $\mathbb{N}$ or the integers $\mathbb{Z}$ as the index set I. There is at least one important example in this course (viz., the Fourier transform) where the real line $\mathbb{R}$ is used as the set I.

A student who has not seen before the notion of the analysis map (i.e., essentially every student in 515) may be surprised by the attention we give above to the index set I. "We may choose, perhaps, very carefully the linear functionals that go into Λ, but the way we index them is a matter of formality", right? Wrong. One of the major arts of data representation is to organize the linear functionals correctly. They usually come in "groups" and "units", and we would like the index set to reflect that organization. □

The key in successful representation is to take a *cohesive* set of linear functionals. Same as the army analog: if you just pick random soldiers and put them together, you might call that an army, but that army cannot fight. You need soldiers that many of them are of "similar skills", and each of them has a well-defined role.

The time representation. The simplest of all representations is the time one. In this case $I := \mathbb{R}$, and Λ is the collection λ_t, $t \in \mathbb{R}$, where λ_t is the point-evaluation at t. The analysis map then provides us with the representation

$$\Lambda^* f = (f(t))_{t \in \mathbb{R}}.$$

For many of us, this looks like a triviality: we just took a function a found a complicated way to recapture something that is identical to the function. The truth is somewhat different: we are used to be "given a function" via its time representation, i.e., via its standard definition as a map from $\mathbb{R}$ to $\mathbb{R}$. The "time representation" is simply the way most of us think about functions in the first place. □

One can easily realize the derivative function f' of f via a suitable selection of another class of linear functionals (which one?)

Definition: sampling. We say that the analysis map Λ^* is a sampling map if each of the linear functionals in Λ is local averaging. □

Regular sampling. We highlight in the sequel the basic assumption that *time is invariant*. If we find a need to do, say, some local averaging over some interval $[a, b]$, then we have an equivalent need to do the same local averaging over any translation of that interval $[a + t, b + t]$, $t \in \mathbb{R}$. For practical purposes, we discretize the above as follows:
First, we choose some local averaging function g, and define the linear functional

$$\lambda_0 : f \mapsto \int_{\mathbb{R}} f(t)g(t)\, dt.$$

Then, for every *integer* j, we define λ_j to be local averaging with respect to the shift of g by i:

$$t \mapsto g(t - j).$$

We denote that shifted function by $g(\cdot - j)$. The analysis map associated with the collection $\Lambda := \{\lambda_j : \ j \in \mathbb{Z}\}$ is called **regular sampling**. Note that every regular sampling is also sampling. A *signal* is the image of a function under regular sampling. □

Regular sampling does not change in an essential way the time representation of a function: we just average the values of the function over many small intervals. In a sense, it is a practical way to record a function using some sensing tool (a sensor, almost always, records local averages of a function).

(4:) Basic assumption: time is invariant.

Let us move ahead with the theory. If time is invariant, and we have picked some function g together with the associated linear functional λ_g then we ought to accept also all the linear functionals which are induced by all the possible *translations* of g.

We define the **translation** operation

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

$E^t(g)$ is then a **translate** of g. We also write $E^t g$, and also $g(\cdot - t)$ for that translate.

The analysis map associated with a function g and all it translates is known as **convolution** and is defined by

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

In fact, it is not g and its translates that appear here but a "flip" of g around the vertical axis. One explanation for that is the desire to get nice properties for the binary operation $f * g$:

Properties of convolution.

(a) commutative: $f * g = g * f$,

(b) associative: $(f * g) * h = f * (g * h)$,
(c) distributive over addition: $f * (g + h) = (f * g) + (f * h)$.

Example. Let B_1 be the support function of the interval $[0, 1]$ (also known as a B-spline of order 1):

$$B_1(t) := \begin{cases} 1, & 0 \le t \le 1, \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

Let f by some function. Then,

$$B_1 * f(t) = \int_{t-1}^{t} f(u)\, du.$$

(verify that!). Thus, the convolution product $B_1 * f$ averages the values of f. □

Example. Choose $f = B_1$ in the previous example. Then $B_2 := B_1 * B_1$ is the hat function, also known as the B-spline of order 2. □

Our next task is to introduce the *frequency representation*. Recall the definition of the complex exponential function:

$$u \mapsto e^{iu} = \cos(u) + i \sin(u).$$

The function

$$e_{i\omega} : t \mapsto e^{i\omega t}$$

is known as **the exponential function with frequency** ω. It is periodic with period $2\pi/\omega$.

The action of multiplying a function f by an exponential function $e_{i\omega}$ is called **modulation**. The frequency representation is obtained by choosing the linear functionals in the analysis map to be exponentials with certain frequencies. This representation is important enough to warrant the opening of the new part.

Before we depart from Part 1, we want to highlight two important operators that were introduced in this section, and to add a third one.

Operation number 1: convolution. It transforms the function f into the convolution product $f * g$.

Operation number 2: modulation. It transforms a function f into the pointwise product $f e_{i\omega}$.

Operation number 3: dilation. Given $a \in \mathbb{R}$, the dilation $\mathcal{D}_a f$ of the function f is a new function defined as

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

We will refer to $\mathcal{D}_a f$ is the a-dilation of f.

Part 2. Introduction: Fourier series and orthonormal systems.

The *frequency representation* is centered around the exponential functions. Whatever version of it we choose (and we will describe two of its many versions), the analysis map of the frequency representation is comprised of exponentials. We learned before to think about a linear functional as a "feature detector". Linear functionals such as local averaging capture "features" which we may be able to visualize and understand in terms of the time representation of the function. The exponentials, being periodic and of global support, are anything but local in time. The "feature" that the exponential linear functional $e_{i\omega}$ captures is not of a kind that had a counterpart in the time representation. Thus, the frequency representation will reveal in our original functions features that the time representation is unable to easily identify. In turn, some properties of the function which are obvious from the time representation (e.g., the interval support of a compactly supported f) become almost "invisible" on the frequency domain.

Fourier series

In the theory discussed in this part, we analyse a space of functions which is known as $L_2(\mathbb{T})$. A function $f : \mathbb{R} \to \mathbb{C}$ belongs to that space if it satisfies the following two conditions:

(1) It is 2π-periodic:

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

(2) It is of finite energy:

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

is finite.

The quantity

$$\|f\|_{L_2}$$

is referred to as **the L_2 (read: "L-two") norm of f.**

The first condition in the above definition is very restrictive. The second one is actually very mild: functions that are continuous definitely satisfy it. Furthermore, functions that have finitely many jump-discontinuities (on the interval $[-\pi, \pi]$) also satisfy it. A guiding example is as follows: the (restriction to $[-\pi, \pi]$ of the) function $|t|^{-1/2}$ is not in L_2. However, $|t|^{-1/3}$ is in L_2 (the actual definition of these functions at the origin is immaterial).

The above norm is associated with the following inner product:

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

The connection is:

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

A norm, in general, is a number that attempts to measure the "size" of a function (there is a more precise mathematical definition to that, and we may come back to this point later on). Most examples of norms *cannot* be associated with an inner-product as above. This makes the L_2-norm very special. We say that $L_2(\mathbb{T})$ is *an inner product space*.

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

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

Note that we are using the same notation, $\|\cdot\|$ for two different norms: the L_2-one and the ℓ_2-one. Usually, it should not cause any confusion: if we write $\|f\|$, you need only to check whether f is a 2π-periodic function or a sequence defined on the integers in order to choose the right norm. Sometime we stress an underlying norm by writing

$$\|x\|_{\ell_2}.$$

The ℓ_2-space is also an inner-product space, with the inner product being

$$\langle x, y\rangle := \langle x, y\rangle := \sum_{n=-\infty}^{\infty} x(n)\overline{y(n)}, \quad x, y \in \ell_2.$$

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.,

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

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.,

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

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 encounter 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.

Functions in the space $L_2(\mathbb{T})$ can be very wild. They might be unbounded, and have an abundant of "bad points": blow-up points, jump-discontinuities, cusps and other, more hidden, bad points. The worse the function is, the harder it is to capture it easily. For example, it might be hard to analyse such function with the aid of only a few values of it.

We try to understand the "goodness" of a function via different methods. The first and most common one is the notion of "smoothness" which amounts to the number of times that we are able to differentiate that function.

Definition: smoothness. Let $f \in L_2(\mathbb{T})$, and let k be a non-negative integer. We say that f is k-times differentiable (or is "differentiable of order k") if:
(a) The $k-1$ derivative of f exists everywhere and is continuous.
(b) The kth derivative of the function exists everywhere with the exception, perhaps, of finitely many points. Moreover, $f^{(k)} \in L_2(\mathbb{T})$. □

Unfortunately, we "cut a few corners" in the above definition: we demanded a bit too much in our definition.

Theorem. *Let k be a non-negative 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.*

While the theorem is entirely correct, you should be warned that the smoothness notion it uses is not entirely the same as ours (it is a bit more forgiving).

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

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

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

Discussion. The Fourier series representation provides, for *every* function f in the space $L_2(\mathbb{T})$, a representation as a convergent series:

$$f = \sum_{n=-\infty}^{\infty} \widehat{f}(n) e_{in}.$$

From a practical point of view, this is "general nonsense": first, the space $L_2(\mathbb{T})$ is very large, and most of the functions in that space are of no interest to us. Second, a representation of functions as infinite series is nice in theory, but in practice, we will have to ignore most of the summands in the series. In fact, we would like to get a good approximation to f by summing a small number of terms. Specifically, we may define, for every positive integer N,

$$f_N := \sum_{n=-N}^{N} \widehat{f}(n) e_{in},$$

and would like to use f_N (for some small value of N) instead of f. Our success in doing that hinges on the size of the error

$$g_N := f - f_N = \sum_{|n|>N} \widehat{f}(n) e_{in}.$$

Now, the Fourier coefficients of g_N are

$$\widehat{g_N}(n) := \begin{cases} 0, & |n| \le N, \\ \widehat{f}(n), & \text{otherwise.} \end{cases}$$

(This is due to the fact that, given any sequence $x \in \ell_2$, that series $\sum_{n\in\mathbb{Z}} x(n)e_{in}$ converges in a function in $L_2(\mathbb{T})$ whose Fourier coefficients are the given values of x.)

Now, by Parseval's formula,

$$\|g_N\|^2 = \sum_{n\in\mathbb{Z}} |\widehat{g_N}(n)|^2 = \sum_{|n|>N} |\widehat{f}(n)|^2.$$

If the Fourier coefficients of f decay fast to 0 as $n \to \pm\infty$, then we expect, already for small values of N, that $\|g_N\|$ is small. This is the case when the function f is known to be smooth, and the smoother the better!

All that sound good: in science we are usually interested in "good" function. If we classify "good" by "smoothness", then we a good deal from the Fourier representation: smooth functions are represented quite accurately using only few terms in the Fourier expansion. However, a glimpse at the example that precedes this discussion reveals the main shortcoming of the Fourier series: the function f in the example has only two bad points, and is very nice elsewhere. Altogether, it is a very simple function, but in terms of smoothness it is not smooth at all (not even continuous). 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). So, it is hard to reconstruct this function from its Fourier coefficients, i.e., we need a large N in order for f_N to be close enough to f.

There is another difficulty that the example reveals. While one can immediately conclude from the slow decay of the Fourier coefficients 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, or how frequently they occur.

In summary, the discussion here exposes two shortcomings of the frequency representation:

(i) It provides good effective representation to smooth functions, but there are many "nice" functions that are not smooth, and for those the representation might be quite deficient (= we need to sum many terms before getting a good "resolution" of the function.)

(ii) It is not local in time: by inspecting the rate of decay of the coefficients we might understand the exact smoothness of the function, but we won't be able to know where the bad points of the function 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

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

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 a 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 allow 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 where the function can be examined: 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.,

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

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:

$$\widehat{f * g} = \widehat{f}\widehat{g}. \tag{4}$$

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 (4) implies, almost immediately, that

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

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

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

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. □

notes will be expanded here: we will add discussion as follows: we will take a nice, smooth, non-negative compactly supported function, and will see how the Fourier transform fails to account for any translation that we apply to the function, while it nicely captures any modulation that we apply to the function.

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 5. *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 6. *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 system that satisfies the complete reconstruction property 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, MRA, and refinable functions

(The beginning of this section repeats the discussion in Part 4, and is included since we skip in class Part 4).

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. This is the option we choose here. It leads us to the notion of *wavelet systems*.

Vanishing moments. Before we continue with our main topic, constructing a wavelet system, we digress in order to analyse the important notion of *vanishing moments*.

Definition. Let ψ be a compactly supported function. We say that ψ has m **vanishing moments** (with m some positive integer) if, for every polynomial p of degree $< m$,

$$\langle p, \psi \rangle = \int_{-\infty}^{\infty} p(t)\, \psi(t)\, dt = 0.$$

Note that the polynomial p is not in our class $L_2(\mathbb{R})$ (why?). Nonetheless, the inner product is well-defined (why?). Also, we are assuming ψ to be real valued. If ψ is complex-valued, a conjugation should be added above (that would not change the definition, though. (why?).) □

Proposition 7. *A compactly supported function ψ has m vanishing moments if and only if*

$$(\widehat{\psi})^{(l)}(0) = 0, \quad 0 \le l < m,$$

i.e., the Fourier transform of ψ has mth order zero at the origin.

Proof: (main idea) Since integration is linear, the vanishing moments property amounts to

$$\int t^l \psi(t)\, dt = 0, \quad 0 \le l < m.$$

Now,

$$\widehat{\psi}(\omega) = \int \psi(t) e^{-i\omega t}\, dt.$$

Differentiating both sides of this equation l times gives

$$\widehat{\psi}^{(l)}(\omega) = \int \psi(t)(-it)^l e^{-i\omega t}\, dt.$$

(We exchanged the order of differentiation and integration , and then found that the argument ω appears only in the exponential. We then used the fact that $(e^{at})' = ae^{at}$, which is valid for complex exponentials as well).

Evaluating the last identity at $\omega = 0$, gives

$$\widehat{\psi}^{(l)}(0) = (-i)^l \int \psi(t) t^l\, dt.$$

From this, it is easy to obtain the claim. □

Theorem 8. *Let ψ be a function supported in the interval $[a, b]$ and having m vanishing moments. Suppose that $f \in L_2(\mathbb{R})$ has m continuous derivatives on the interval $[a, b]$. Denote*

$$C(f, a, b) := \max\{|f^{(m)}(t)| : t \in [a, b]\},$$

and assume that $\|\psi\| \le 1$. Then

$$|\langle f, \psi \rangle| \le C\, C(f, a, b)(b - a)^{m+1/2}.$$

Proof: Let p be the Taylor expansion of f of order m around the point a, i.e., p is a polynomial of degree $< m$. Writing

$$E(t) := \begin{cases} f(t) - p(t), & t \in [a, b], \\ 0, & \text{otherwise}, \end{cases}$$

it is known that, for every $t \in [a,b]$, there exists $c_t \in (a,t)$ such that

$$E(t) = \frac{f^{(m)}(c_t)}{m!}(t-a)^m.$$

Thus, by our assumptions, $|E(t)| \leq \frac{C(f,a,b)(b-a)^m}{m!} =: K$.

Now, due to the vanishing moments of ψ,

$$\langle f, \psi \rangle = \langle p, \psi \rangle + \langle E, \psi \rangle = \langle E, \psi \rangle.$$

(Note that our decision to define E to be 0 outside $[a,b]$ affects nothing, since ψ already vanishes outside $[a,b]$. I.e., $p+E$ equals f only on $[a,b]$, and that's all we need). Therefore,

$$|\langle f, \psi \rangle| = |\langle E, \psi \rangle| \leq \|E\|\|\psi\| \leq \|E\|.$$

Finally,

$$\|E\| = \sqrt{\int_a^b |E(t)|^2\, dt} \leq \sqrt{\int_a^b K^2 dt} = K(a-b)^{1/2} = \frac{C(f,a,b)}{m!}(b-a)^{m+1/2}.$$

□

Remark. We have forgone listing the constant C as $1/m!$ since better constants are available. For example, we could do the Taylor expansion around the midpoint of the interval.

Discussion. If the mother wavelet ψ is supported in $[\alpha, \beta]$, then the wavelet $\psi_{j,k}$ is supported in $[2^{-j}(\alpha+k), 2^{-j}(\beta+k)]$. The length of its support is, therefore, $(\beta-\alpha)/2^j$. Also, $\|\psi_{j,k}\| = \|\psi\|$ (why?). Thus, if f has m continuous derivatives in some fixed interval $[a,b]$, then, for every wavelet $\psi_{j,k}$ whose entire support lies in $[a,b]$, we will have

$$|\langle \psi_{j,k}, f \rangle| \leq C\, C(f,a,b)(\alpha-\beta)^{m+1/2} 2^{-j(m+1/2)}.$$

Note that each time we increase j by 1, the bound above shrinks by a factor of $2^{m+1/2}$. Also note that extra differentiability in f does not help here, since we are limited by the number of vanishing moments in ψ. Also, additional vanishing moments in ψ are not going to help either, unless f has additional derivatives. So, the basic rule is:

Basic rule. The expected decay rate of the wavelet coefficients is determined by the smaller of (i) the number of vanishing moments of the wavelet, and (ii) the smoothness of f on the support of the corresponding wavelet. □

This is now a leftover. I am leaving it in just in case there is still some value in it. will be expanded here. we will add a discussion how we construct a good window function ψ for the wavlet system. The discussion will say that we try achieve three goals:

(1) The window function should be local in time. This is achieved by selecting a function ψ with compact support. Ideally, the support is the interval $[0,1]$ (for reasons that we do not elaborate on now it cannot be smaller than that. Usually the support will be larger.)

(2) The window function should be local in frequency, i.e., should be smooth.

(3) The inner product $\langle f,\psi\rangle$ should measure the "roughness of f": if f is a very nice function (on the support of ψ: $\langle f,\psi\rangle$ does not depend at all on the behavior of ψ outside the support of f. why?), we want $\langle f,\psi\rangle$ to be small.

The last point requires some elaboration. Let us agree that a linear polynomial at our epitome of anything that is "nice and predictable" (after all, if you know two values on the graph of the linear polynomial, you find all the other values by simple averaging. So it is a very predictable function). Now, suppose that ψ is supported on an interval $[a,b]$. We look at f on that interval, and we try to fit (on that interval only) the function f a linear polynomial p:

$$f = p + (f - p).$$

Suppose that we managed to do very well, i.e., we found p such that the error $f-p$ is a very small on $[a,b]$. In this case, it is reasonable to think about f is " very predictable" (i.e., smooth) on the interval. If ψ measures "roughness" (and is supported on $[a,b]$) then $\langle f,\psi\rangle$ may better be small, reflecting thereby the "goodness" of f on $[a,b]$. An easy way to do that is to require that

$$\langle p,\psi\rangle = 0, \quad \text{for every linear polynomial.} \tag{9}$$

Note that, by taking $p(t)=1$ in (9) we obtain the requirement

$$\widehat{\psi}(0) = \int_{-\infty}^{\infty} \psi(t)\,dt = \langle 1,\psi\rangle = 0.$$

In fact, it can be shown (and is not very hard) that (9) is equivalent to

$$\widehat{\psi}(0) = \widehat{\psi}'(0) = 0. \tag{10}$$

Now, if we managed to find ψ that satisfies (9), then, for any linear polynomial p

$$\langle f,\psi\rangle = \langle p + (f-p),\psi\rangle = \langle p,\psi\rangle + \langle f-p,\psi\rangle = \langle f-p,\psi\rangle.$$

So, if ψ is supported on $[a,b]$, and there is, on $[a,b]$, a good fit to f by some linear p, then we are guaranteed that $\langle f,\psi\rangle$ is small, without a need to ever find p!!

If we consider 'roughness' as the "inability to find a good fit to f on small intervals by linear polynomials" (a notion which is closely related to failing to differentiate f even once, hence is a very good way to think about roughness), then all the above tells that that we better require of ψ to satisfy (10).

Thus,,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 $j \geq 0$:*

$$\langle \psi_{0,0}, \psi_{j,k} \rangle = \begin{cases} 1, & j = k = 0, \\ 0, & \textit{otherwise.} \end{cases} \tag{11}$$

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 (11) 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 (11), since $\mathcal{D}_{2^{j'-j}} E^{k' - 2^{j'-j} k} \psi = \psi_{j'-j, k' - 2^{j'-j} k}$. $\square$

Remark. Another variant of this result is as follows: W_ψ is orthonormal if and only if the following conditions are valid, for every $j \leq 0$ and every $k \in \mathbb{Z}$:

$$\langle \psi_{j,0}, \psi_{0,k} \rangle = \begin{cases} 1, & k = j = 0, \\ 0, & \text{otherwise}, \end{cases}$$

We organise the wavelets in W_Ψ in *scales* or *layers* or *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 wider...

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:

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

(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 Shannon wavelet system tiles 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:

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

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 strictly '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: We would like to construct wavelet systems that satisfy the following conditions:

(1) The system is local in time. We can achieve that by making sure that the mother wavelets in our system have (short) compact support.

(2) The system is local in frequency. There are two complementary aspects here. One is the requirement that the Fourier transform of each mother wavelet decays fast at ∞, or equivalently, that each mother wavelet be smooth. As alluded to above, we also require the Fourier transform of each mother wavelet to "stay away of the origin". We will come back to that notion later.

(3) The analysis/decomposition should be inverted by a corresponding synthesis step. For example, we may require the system to satisfy the perfect reconstruction formula:

$$f = \sum_{j,k\in\mathbb{Z}} \sum_{\psi\in\Psi} \langle f, \psi_{j,k}\rangle \, \psi_{j,k}, \quad \forall f \in L_2(\mathbb{R}).$$

Recall that a complete orthonormal W_Ψ would have this perfect reconstruction property but not vice versa.

(4) The decomposition step as well as the reconstruction step should be implemented by fast algorithms.

Sometime, we also add another requirement: 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 12. Let $\phi \in L_2$ be a given function. We say that the function ϕ is **refinable** if we can write $\mathcal{D}_{1/2}\phi$ as a linear combination of $E(\phi)$. □

We will generalize 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

$$\phi(t/2) = \phi(t) + \phi(t-1). \tag{13}$$

We can write this relation also as

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

□

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

$$\phi_{-1,0} = \sum_{k\in\mathbb{Z}} c(k)\phi_{0,k}, \tag{14}$$

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. With this normalization, the refinement equation (14) reads as

$$\phi_{-1,0} = \sqrt{2}\sum_{k\in\mathbb{Z}} h(k)\phi_{0,k}. \tag{15}$$

We note that there is a simple way to check whether the sequence h is normalized correctly: when normalized properly, its values should sum up to 1.

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:

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

Clearly, the refinability condition can be restated as

$$\mathcal{D}_{1/2}\phi \in V_0(\phi).$$

Set now

$$V_{-1}(\phi) := \{\mathcal{D}_{1/2} f : \ f \in V_0(\phi)\},$$

Thus, $V_{-1}(\phi)$ is obtained by applying dilation by 1/2 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 *even-shifts* of the dilated ('stretched') function $\phi_{-1,0}$. One can check that the refinability amounts to the statement that

$$V_{-1}(\phi) \subset V_0(\phi).$$

The definition of $V_0(\phi)$ is a bit vague, since $E(\phi)$ contains infinitely many functions, and the Linear Algebra sense of `span` usually discusses the span of *finitely many* vectors/functions.

At this point, it is more important to pay attention to the nature of $V_0(\phi)$ than to the above nuance. For example, if $\phi = B_1$, then $V_0(\phi)$ consists of piecewise-constants with (possible) integer breakpoints. $V_{-1}(\phi)$ consists of piecewise-constants with (possible) break-points at the even integers (why?). Thus, indeed, the latter is a subspace of the former.

If $\phi = B_2$, then $V_0(\phi)$ consists of continuous piecewise-linear functions with integer breakpoints. $V_{-1}(\phi)$ consists then of continuous piecewise-linears with (possible) breakpoints at the even integers. Again, the the latter is a subspace of the former. This means that B_2 is refinable, too. Let's discuss this example in detail.

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

$$\phi_{-1,0} = \sqrt{2}(\frac{1}{4}\phi_{0,-1} + \frac{1}{2}\phi_{0,0} + \frac{1}{4}\phi_{0,1}).$$

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, more general and sometimes more convenient, way to understand the notion of refinability: viewing the refinability condition as connecting the Fourier transforms of ϕ and $\mathcal{D}_{1/2}\phi$. Let's pause momentarily and make some observation before taking that new route.

Suppose that f is writable as a finite linear combination of $E(\phi)$:

$$f = \sum_{k=-N}^{N} c(k) E^k \phi.$$

Applying the Fourier transform to both sides of the above equation, and using some of the known properties of that transform we get

$$\widehat{f} = \widehat{\phi} \sum_{k=-N}^{N} c(k) e_{-ik}.$$

The sum $\sum_{k=-N}^{N} c(k)e_{-ik}$ is clearly 2π-periodic. It is actually a very special type of a 2π-periodic function since it is a *finite* combination of the periodic exponentials. Such function is known as a **trigonometric polynomial**. The exponentials themselves (with integer frequency) are trivially trig. polynomials. More interesting examples are cos and sin. Indeed (check!):

$$\cos(\omega) = \frac{e^{i\omega} + e^{-i\omega}}{2} \quad \sin(\omega) = \frac{e^{i\omega} - e^{-i\omega}}{2i}.$$

As a summary of the above, one concludes that, if $f \in L_2(\mathbb{R})$ and

$$\widehat{f} = H\widehat{\phi},$$

with H some trig. polynomial, then $f \in V_0(\phi)$. This allows us to provide the following (complete and correct!) definition of $V_0(\phi)$:

Definition 16. *Let ϕ be some $L_2(\mathbb{R})$ function. We define $V_0(\phi)$ as the collection of all functions $f \in L_2(\mathbb{R})$ that satisfy an equation*

$$\widehat{f} = H\widehat{\phi}$$

with H some 2π-periodic function. □

In view of the above, we can recast the refinability condition

$$\mathcal{D}_{1/2}\phi \in V_0(\phi)$$

on the Fourier domain. Since $\widehat{\mathcal{D}_{1/2}\phi} = \mathcal{D}_2\widehat{\phi}$, we obtain that ϕ is refinable if there exists 2π-periodic H_0 such that

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

(we absorbed the normalization constant in $\mathcal{D}_2\widehat{\phi}$ into the definition of H_0). Thus:

Complete and Correct Definition of Refinability 17. *A function ϕ is said to be refinable if its Fourier transform satisfies an equality of the form*

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

with H_0 a 2π-periodic function. From now on, we will require additionally that $H_0(0) = 1$. □

A simple exercise shows that if

$$\phi_{-1,0} = \sqrt{2} \sum_{k=-N}^{N} h(k)\phi_{0,k}$$

then

$$\widehat{\phi}(2\omega) = H_0(\omega)\widehat{\phi}(\omega), \quad H_0 = \sum_{k=-N}^{N} h(k)e_{-ik},$$

i.e., H_0 is the Fourier series of sequence h. The function H_0 is also called the mask of the refinable ϕ, or the **symbol** of ϕ.

As a first example of the frequency interpretation of refinability, apply the Fourier transform to equation (13). Then (after some simple calculation) we get that (for $\phi := B_1$),

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

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

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

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_0 is defined by the relation

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

Using the correct trigonometric identity (which one?) one easily gets that $H_0(\omega) = \cos^2(\omega/2)$. □

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_0 be the (2π-periodic extension) of the support function of the interval $[-\pi/2, \pi/2]$. Then, obviously, ϕ is refinable (according to definition (17)) with mask H_0. □

Part 6: MRA and the unitary extension principle

The MRA approach. Start with a refinable ϕ. Choose the mother wavelets Ψ such that each $\psi \in \Psi$ satisfies $\mathcal{D}_{1/2}\psi \in V_0(\phi)$.

For example, if $\phi = B_1$, then we choose each $\mathcal{D}_{1/2}\psi$ to be some (carefully selected) piecewise-constant with integer breakpoints. This means that ψ itself is piecewise-constant with half-integer breakpoints (why?). Of course, this shows that Haar is derived from the MRA of B_1. □

Discussion. The smoothness of the mother wavelets will be determined by the smoothness of the refinable function. The MRA setup guarantees fast algorithms (as we will see later). In order to build mother wavelet that have small support (local in time), we need to choose a refinable function with short, compact, support, and we need also to make sure that only a few of the shifts of this ϕ are involved in the construction of the mother wavelets (why?). And how are we going to achieve the prefect reconstruction property (or, even better, to construct an orthonormal W_Ψ)? □

In order to address this latter question, we formalize it as follows. First, we are given a refinable function ϕ, namely, we are given a 2π-periodic function H_0 that satisfies

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

In addition we assume throughout that

$$H_0(0) = 1$$

(refinable functions that violate this conditions are pathological). We want to choose mother wavelets $\psi_1, \psi_2, \ldots, \psi_r$ such that $\mathcal{D}_{1/2}\psi_i \in V_0(\phi)$, $i = 1, \ldots, r$. Looking at the definition (16) of $V_0(\phi)$, we conclude that the condition $\mathcal{D}_{1/2}\psi_i \in V_0(\phi)$ implies (how?) that there exists a 2π-periodic function H_i such that

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

Note that the mother wavelets (hence the entire wavelet system) are determined by ϕ and $(H_i)_{i=1}^r$.

So, the MRA construction of wavelet systems can be described as a 2-step approach:

Step I: Pick a refinable function ϕ of desired properties. Find its refinement mask H_0. ϕ is either chosen from the known refinable functions (e.g., a B-spline), or, sometimes, we will need to construct it so that it has some especially tailored properties.

Step II: Choose the mother wavelet $\psi_1, \ldots, \psi_r$, or, in other words, choose the wavelet masks $H_1, \ldots, H_r$.

The ultimate goal in the MRA construction of wavelets 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:

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

Even more ambitiously, we might want the system to be complete and orthonormal.

(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 $V_0(\phi)$ will be smooth, and the mother wavelets will inherit this smoothness. 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 masks (H_i) to be trigonometric polynomials (show that!).

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

Definition 18. *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:*

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

□

Theorem 19 (The unitary extension principle (=UEP)). *Let ϕ be refinable with mask H_0, $H_0(0) = 1$. Let $H_1, \ldots, H_r$ be 2π-periodic functions. Assume that the following two conditions hold for every $\omega \in [-\pi, \pi]$:*

$$|H_0(\omega)|^2 + |H_1(\omega)|^2 + \ldots + |H_r(\omega)|^2 = 1, \tag{20}$$

and

$$H_0(\omega)\overline{H_0(\omega+\pi)} + H_1(\omega)\overline{H_1(\omega+\pi)} + \ldots + H_r(\omega)\overline{H_r(\omega+\pi)} = 0. \tag{21}$$

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

Example. We examine the binomial expansion of

$$(\cos^2(\omega/2) + \sin^2(\omega/2))^2, \tag{22}$$

and take H_m, $m = 0, 1, 2$ to be the square root of the $(m+1)$st 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 (20) is obviously valid here, and condition (21) 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 $h_2(0) = \frac{1}{2}$, hence

$$\psi_2(t) = 2(h_2(1)\phi(2t-1) + h_2(0)\phi(2t) + h_2(1)\phi(2t+1)) = 1 - \frac{1}{2}(\phi(2t-1) + \phi(2t+1)).$$

The above example generalizes to higher order B-spline. One just needs to use a higher power in (22). Note that the number of wavelet increases together with that power. □

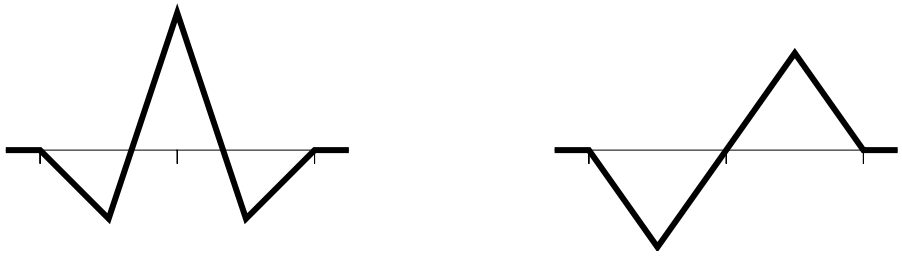

Figure. The two mother wavelets that generate a piecewise-linear tight frame (ψ_1 is on the right).

Part 7: constructing orthonormal wavelet systems

A complete orthonormal system is a tight frame but not vice versa. The performance of the some of the known wavelet algorithms is better understood for orthonormal systems and less for the more general tight frames. Thus, one may insist on constructing an orthonormal system and not only a tight frame. In this regard, it is useful to note the following:

Fact. *If W_Ψ is a complete and orthonormal wavelet system, then Ψ is a singleton, i.e., the system employs a single mother wavelet.* □

We would like then to focus on UEP constructions that are built on a single mother wavelet. First, let us interpret the UEP as follows: let $H_0, \ldots, H_r$ be the masks involved in the UEP construction. We introduce a matrix with $r+1$ rows and 2 columns as follows:

$$\begin{pmatrix} H_0(\omega) & H_0(\omega+\pi) \\ H_1(\omega) & H_1(\omega+\pi) \\ . & . \\ . & . \\ H_r(\omega) & H_r(\omega+\pi) \end{pmatrix}.$$

Denote the columns of this matrix by $H(\omega)$ and $H(\omega+\pi)$. Then the UEP simply requires these two vectors to be *orthonormal* for every fixed ω.

Now, suppose that we require $r = 1$. Then the above matrix is square (2×2), and the orthonormality of its columns then implies that it is a unitary matrix, and in particular that the rows of that matrix are orthonormal, too. If we think about H_0 as given, and about H_1 is being sought-for, then the above discussion shows that we cannot succeed (in using the UEP with a single mother wavelet) unless H_0 is a CQF:

Definition 23. Let H_0 be a 2π-periodic function. We say that H_0 is a CQF (Conjugate Quadrature Filter) if $H_0(0) = 1$ and

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

Example. Let

$$H_0(\omega) = e^{i\omega/2}\cos(\omega/2) = \frac{1+e^{i\omega}}{2}.$$

Then $|H_0(\omega)|^2 = \cos^2(\omega/2)$, and it follows then that H_0 satisfies the CQF condition. Note that this particular H_0 is the mask of B_1. □

So, suppose that H_0 is a CQF. Can we then find H_1 so that the UEP conditions are satisfied? The answer is YES.

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 (20) is trivially satisfied. However, this is too crude for the satisfaction of (21). Instead, we define the unique wavelet mask to be

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

Then (20) and (21) 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.

At this point, the CQF construction is still very fishy. First, in our previous constructions we started with a refinable function ϕ whose mask is H_0. Here we started with H_0 without paying attention to the question whether there is a refinable ϕ with such mask. This difficulty is settled in the following result:

Theorem. *Let H_0 be a trigonometric polynomial, and assume that H_0 satisfies the CQF condition.*
Then there exists a compactly supported function $\phi \in L_2$ which is refinable with mask H_0.

Next, we realize that the CQF construction is thus a special case of the UEP, and hence the resulting wavelet system is a tight frame. However, we want to obtain an *orthonormal system*. The next fact helps us in this regard:

Theorem 24. *Let ϕ be a compactly supported refinable function whose shifts $E(\phi)$ are orthonormal. Then:*

(i) The mask H_0 of ϕ satisfies the CQF condition.

(ii) The CQF construction yields a wavelet system W_ψ which is complete and orthonormal.

We note that (i) above cannot be reversed: there are refinable functions whose mask is a CQF, but whose shifts are *not* orthonormal. In fact, we have the following precise statement:

Theorem 25. *Let H_0 be a trigonometric polynomial, and assume that H_0 satisfies the CQF condition. Let ϕ be a corresponding refinable function, and assume that $\widehat{\phi}(0) = 1$. Then 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 examples. 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_0, once we know that $\widehat{\phi}$ does not have a 2π-periodic zero.

As an example, take

$$H_0(\omega) := \frac{1 + e^{-3i\omega}}{2}.$$

Then H_0 is a trigonometric polynomial, $H_0(0) = 1$ and H_0 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?) Indeed, the wavelet system constructed by applying the CQF unitary extension principle to this H_0 is a tight frame (as it should), but is not an orthonormal system. □

Summary: how to construct an orthonormal wavelet system? Start with a refinable ϕ whose shifts $E(\phi)$ are orthonormal. In fact, since such functions are not really around, start with a CQF H_0. Then apply the CQF construction. If you can show that you cannot satisfy (2) in Theorem 25, then you must satisfy (1) of that theorem, and you made it: W_ψ is complete and orthonormal.

Fortunately, among the two possibilities in Theorem 25, (1) is abundant and (2) is rare, so a generic CQF construction yields indeed an ortho wavelet basis.

Before we move on, it is worthwhile to clarify one point: suppose that we select a trig. polynomial H_0 which is *not* a CQF. Is it still possible to find a refinable ϕ whose mask is H_0? and if so, then how do we find such a function?

For that, let us look at the infinite product

$$\nu(\omega) := \prod_{j=1}^{\infty} H_0(\omega/2^j). \tag{26}$$

Theorem 27. *Let H_0 be a trigonometric polynomial and assume that $H(0) = 1$. Then the infinite product (26) converges everywhere (to a very smooth function).*

The theorem is insufficient: it does not tell us that the limit ν is the Fourier transform of some $\phi \in L_2$. It turns out that this is a much harder problem, which we will settle once we deal with the issue of the *smoothness* of refinable functions.

Construction of Daubechies' refinable functions.

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

$$(\cos^2(\omega/2) + \sin^2(\omega/2))^{2k-1}, \tag{28}$$

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 $k = 2$,

$$T(\omega) = \cos^6(\omega/2) + 3\cos^4(\omega/2)\sin^2(\omega/2). \tag{29}$$

$T(\omega)$ 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 (28), and hence

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

The Féjér-Riesz Lemma 30. *Let T be a trigonometric polynomial which is non-negative everywhere. Then there exists a trigonometric polynomial H_0 such that*

$$T(\omega) = |H_0(\omega)|^2.$$

Applying this lemma, we obtain a trigonometric polynomial, H_0, such that $|H_0(\omega)|^2 = T(\omega)$. For example, for the case (29) this polynomial turns out to be

$$H_0(\omega) = \cos^2(\omega/2)\left(\frac{1+\sqrt{3}}{2} + \frac{1-\sqrt{3}}{2}e^{i\omega}\right).$$

Note that we can now conclude that (for each value of k) the above H_0 is a CQF (do not forget to verify that $H_0(0) = 1$!) Then, Theorem 25 implies most of next result.

Theorem and Definition. *Given k as above, there exists a corresponding trigonometric polynomial $H_0 := H_{0,k}$, and a function d_k (known as 'Daubechies' refinable function' of order k) such that*

(i) *d_k is refinable with mask H_0.*
(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 $h_{0,k}$ associated with $H_{0,k}$ has exactly $2k$ non-zero coefficients:*

$$H_{0,k}(\omega) = \sum_{m=0}^{2k-1} h_{0,k}(m) e^{-im\omega}.$$

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 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 31. *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_3 (which is supported in an interval on length 5), can be proved to have (barely) one continuous derivative.

In addition to the CQF construction, there are general recipes for constructing other tight wavelet frames based on the unitary extension principle (that employ more than one mother wavelet). For example, note that the construction of the piecewise-linear tight frame (from Part 6) can be generalized to higher order B-splines.

Part 8: 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:

$$\nu_{j,m} := \sqrt{2}(h_m * \nu_{j-1,0})_{\downarrow}. \tag{32}$$

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 (32), 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 9: 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:

$$\nu_{j,m} = \sqrt{2}(h_m * \nu_{j-1,0})_\downarrow, m = 0, \ldots, N. \tag{33}$$

Let's denote by

$$X_{j,m}$$

the Fourier series of $\nu_{j,m}$. With some (but not much) effort, one shows that (33) 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 (20) and (21) 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 relabeling 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 $y := x_\uparrow$, $Y(\omega) = X(2\omega)$.

(2) Conjugation: the switch from $H_m(\omega)$ to $\overline{H_m(\omega)}$ amounts to flipping the entries in the sequence h_m: with h_m^* defined by

$$h_m^*(k) := \overline{h_m}(-k),$$

the Fourier series of h_m^* is $\overline{H_m}$. Note that we also need to apply conjugation to the coefficients, however, in all examples of interest h_m will be real-valued.

(3) Multiplication: The product $\overline{H_m(\omega)} X(2\omega)$ indicates that we need to convolve the underlying sequences, i.e., the sequences h_m^* and $x_\uparrow$.

Summary: The reconstruction step is

$$x_{j-1,0} \longleftarrow \sum_{m=0}^{N} h_m^* * (x_{j,m\uparrow}) X_{j-1,0}(\omega).$$

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

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

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 34. *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 35. *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 11: The transfer operator

As we already know, wavelet systems are constructed via the tool of MRA. This requires a 'good' refinable function to begin with. Desired properties of the refinable function include:

(1) Compact support (since this allows us to construct wavelets of compact support. we simply need to make certain that all the wavelet masks have finite support).

(2) being in L_2. Otherwise, we cannot provide any theory that explains the filter bank algorithm (i.e., the fast wavelet transform).

(3) smoothness. Once you know that the refinable function exists, you'd like it to be smooth. That smoothness is inherited by the wavelets, and the smoothness of the wavelets is necessary for a good frequency localization of the wavelet system.

(4) having orthonormal shifts. That will imply that the unitary extension principle results in wavelet systems that are orthonormal, and not only tight frames.

(5) symmetry (later on)

(6) high approximation order (together with (3) this may enable us to construct wavelets with good frequency localization; we will discuss that issue in the future)

The tool we exploit to this end is called the *transfer operator.* It is a map T that maps 2π-periodic functions to 2π-periodic functions. It is linear, i.e.,

$$T(aX + bY) = aT(X) + bT(Y),$$

for every scalars a, b and every functions X, Y.

Definition: the transfer operator. Let H_0 be a trigonometric polynomial:

$$H_0(\omega) = \sum_{n=0}^{m} h_0(n) e^{-in\omega}.$$

Denote:

$$\mathcal{H}_0 := |H_0|^2.$$

The transfer operator T is then defined as follows: its domain is the space of trigonometric polynomials of degree m:

$$P_m := \text{span}\{e_{in} : n = -m, \ldots, m\}.$$

(recall that e_θ is the exponential $\omega \mapsto e^{\theta\omega}$). Then:

$$T(X)(\omega) = (\mathcal{H}_0 X)(\omega/2) + (\mathcal{H}_0 X)(\omega/2 + \pi).$$

□

It can be shown that

$$T(P_m) \subset P_m.$$

This means that we can talk about eigenvalues and eigenvectors (=eigenfunctions) of T. To recall, given $\mu \in \mathbb{C}$, and $X \in P_m$, we say that μ is an eigenvalue of T associated with the eigenvector X if

$$T(X) = \mu X.$$

Example. Let

$$H_0(\omega) := \frac{1 + e^{-i\omega}}{2}.$$

Then $\mathcal{H}_0(\omega) = \cos^2(\omega/2)$, and the corresponding transfer operator is

$$T(X)(\omega) = \cos^2(\omega/4)\, X(\omega/2) + \sin^2(\omega/4)\, X(\omega/2+\pi).$$

Here $m = 1$, hence the space

$$P_1 = \text{span}\{1, e^{i\omega}, e^{-i\omega}\}$$

is mapped by T into itself (check!). It is immediate that $T(1) = 1$ (i.e., for the constant function $X = 1$, $T(X) = X$). This means that 1 is an eigenvalue associated with the constant eigenfunction $X = 1$.

Moreover, $T(\sin(\omega)) = \frac{\sin(\omega)}{2}$, (check!), which means that $1/2$ is an eigenvalue associated with the eigenfunction $\sin(\omega)$. Finally, $T(\cos(\omega) - 1) = \frac{\cos(\omega)-1}{2}$, which means the a third eigenpair is given by $(\frac{1}{2}, \cos(\omega) - 1)$. Since P_1 is of dimension three, and since we have found three (linearly independent) eigenvectors, we should look no further.

The summary for the transfer operator in this case is as follows: it spectrum $\sigma(T)$ (:= the set of all eigenvalues) is $\{1, 1/2\}$, and the *dominant* eigenvalue 1 is unique and simple. □

Recall that the **spectral radius** of a linear map T (that maps a finite dimensional space into itself) is defined as

$$\rho(T) := \max\{|\mu| : \ \mu \in \sigma(T)\}.$$

Thus, in the above example, $\rho(T) = 1$.

Theorem 36. *Let H_0 be a trigonometric polynomial, $H_0(0) = 1$. Let T be the corresponding transfer operator.*

(i) If there exists a function $\phi \in L_2$ which is refinable with mask H_0 then $\mu = 1$ is an eigenvalue of T, and the corresponding eigenfunction is $X = \sum_{n=-m}^{m} c(n) e_{-in}$, where

$$c(n) := \langle \phi, \phi(\cdot - n) \rangle.$$

Moreover, X is non-negative, i.e.,

$$X(\omega) \geq 0, \quad \forall \omega.$$

(ii) If the spectral radius of T is 1, and if all the eigenvalues of T of modulus 1 are non-defective, then there exists $\phi \in L_2$ which is refinable with mask H_0. In particular, this is the case if 1 is a unique dominant simple eigenvalue of T.

Example: CQF masks. If H_0 is CQF, then

$$\mathcal{H}_0(\omega) + \mathcal{H}_0(\omega + \pi) = 1, \quad \forall \omega.$$

This implies that the transfer operator has in this case the eigenvalue 1 with the constant function being the eigenvector. It is also possible (and not much harder) to prove that in that case $\rho(T) = 1$ (hint: suppose that

$$T(X) = \mu X,$$

for some $|\mu| > 1$. Begin your proof by evaluating the above at the max point of X).

Theorem 37. *Let H_0 be a CQF mask. Then:*
(i) There exists $\phi \in L_2$ which is refinable with mask H_0.
(ii) The shifts of ϕ are orthonormal if and only if 1 is a unique and simple dominant eigenvalue of T.

Discussion. In a previous theorem, we gave a different characterization of the orthonormality issue: if H_0 is a CQF then the shifts of ϕ are orthonormal if and only if ϕ does not have a 2π-periodic zero. It is not clear (albeit not impossible) how to find easily the zeros of $\widehat{\phi}$, given that only H_0 is known. At the same time, finding eigenvalues of linear maps is routine. □

The transfer operator also enables us to find the smoothness of ϕ. Before we state this result, we precede it by the following discussion. Suppose that $\phi \in L_2$. We can then construct a new function with better smoothness (which is refinable, providing that ϕ is one) by convolving ϕ with a B-spline of order n:

$$\varphi := \phi * B_n.$$

In can be shown that φ has all its derivatives up to order n lie in L_2. In short, convolution with B_n increases the smoothness of ϕ by n.

A deeper result is that the converse is also true: if all the derivatives of the compactly supported refinable φ up to exactly order n lie in L_2, then we can 'factor' φ into

$$\varphi = \phi * B_n.$$

While $\phi \in L_2$ and is compactly supported, its first order derivative is not in L_2 (otherwise, φ itself would have been differentiable $n+1$ times).

So, when we analyse the smoothness of φ, we first factor out the largest possible B-spline factor:

$$\varphi = \phi * B_n.$$

At this point, we already know that the $(n+1)$-derivative of φ is not in L_2. While B_n is smooth and nice, 'bad factor' ϕ may decrease the smoothness of product. (So, B_n provides smoothness to convolution, while the bad guy ϕ helps in other ways: e.g., while the shifts of B_n are not orthonormal, the convolution of B_n with ϕ may yield a refinable function ϕ with orthonormal shifts. So, the convolution of B_n with ϕ be accurately thought of as a trade-off, where we lose smoothness, and gain some other desired properties).

In terms of the mask H of φ, a factorization of the form $\varphi = \phi * B_n$ is implied by a factorization

$$H(\omega) = \left(\frac{1+e^{-i\omega}}{2}\right)^n H_0(\omega),$$

with H_0 the mask of the 'bad' ϕ. Equivalently,

$$|H(\omega)|^2 = \cos^{2n}(\omega/2)|H_0(\omega)|^2.$$

The transfer operator enables us to find the 'bad effect' of the second factor H_0. The actual result is as follows.

Theorem 38. *Let H_0 be a given trigonometric polynomial ($H_0(0) = 0$). Let ρ be the spectral radius of the transfer operator associated with H_0. Given any positive n, define*

$$H(\omega) := \left(\frac{1+e^{-i\omega}}{2}\right)^n H_0(\omega).$$

Assume that

$$k := n - \frac{\log_2 \rho}{2} > 0.$$

Then:

(i) There exists a function $\phi \in L_2$ which is refinable with mask H. All the derivatives of ϕ up to order $< k$ exist and lie in L_2.

(ii) If the shifts of ϕ are orthonormal, then the converse is also true: the derivatives of ϕ to any order $\geq k$ is not in L_2.

Example. Let H be the mask of Daubechies' function ϕ of order 2. We know that

$$|H(\omega)|^2 = \cos^4(\omega/2)(\cos^2(\omega/2) + 3\sin^2(\omega/2)).$$

Thus, we define H_0 as the squareroot of

$$\mathcal{H}_0(\omega) = cos^2(\omega/2) + 3\sin^2(\omega/2).$$

(we do not need to find that squareroot!). Let T be the transfer operator associated with $\mathcal{H}_0$, and let ρ be the spectral radius of T. The above theorem says that all the derivatives of ϕ up to any order smaller than

$$2 - \frac{\log_2 \rho}{2}$$

are in L_2. Unfortunately, we can find out that $\rho = 4$ in this case (the eigenfunction is the constant function $X = 1$), which means that

$$2 - \frac{\log_2 \rho}{2} = 1,$$

which then merely implies that only the zero-order derivative of ϕ (i.e., ϕ itself) lies in L_2. Since the shifts of ϕ are orthonormal, the second part of the theorem says that we do not underestimate here the smoothness. I.e., the first derivative of ϕ is not in L_2. □

Part 12: 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, $H_{j,k}$ $(0 \leq j, k \leq 1)$ 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: