

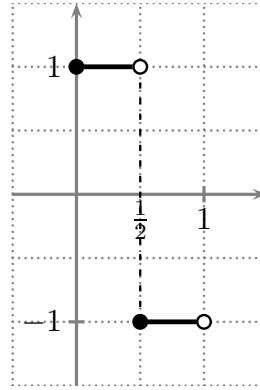
LECTURE 1

INTRODUCTORY MATHEMATICAL IDEAS

The term wavelet, meaning literally ‘little wave’, originated in the early 1980s in its French version ‘ondelette’ in the work of Morlet and some French seismologists.

Prototypical example. Haar wavelet

$$h(x) = \begin{cases} 1, & 0 \leq x < \frac{1}{2}, \\ -1, & \frac{1}{2} \leq x < 1, \\ 0, & \text{elsewhere.} \end{cases}$$



Here the ‘wave’ property is shown in the graph and indicated also by the fact that h has *average zero i.e.*,

$$\int_{-\infty}^{\infty} h(x) dx = \int_0^1 h(x) dx = 0,$$

so the Haar wavelet has some *oscillation* as do trigonometric functions, but unlike such periodic functions as sin and cosine it has *compact support* - it ‘lives’ on the interval $[0, 1)$ and doesn’t oscillate everywhere on $(-\infty, \infty)$ - hence the addition of the word ‘little’.

Much of the impetus for this development, however, came from the realization that the ideas underlying wavelets could be found in many different forms in many different areas of mathematics, science and engineering:

- (I) Harmonic analysis:
- (II) Mathematical physics:
- (III) Electrical engineering, signal processing:
- (IV) Geology:
- (V) Image analysis:

This meant that researchers from many disparate fields were automatically attracted to wavelets. As a result the development of the subject, both theory and applications, has ‘exploded’ since the early 1980s. It is not our intention to cover all these topics, even if

time were available, nor can we hope to describe a lot of the ‘cutting-edge’ applications that have been made of wavelets to these and other areas, but we shall try to develop enough of the ideas that one can at least begin to explore the world of wavelets with an emphasis on signal analysis, admittedly from a mathematical point of view!

We shall use the words *function* and *signal* interchangeably. A function will be a function

$$f = f(x), \quad f = f(x, y)$$

of one or more variables (at most two in these notes), or a sequence

$$a = \{a_n\}_n, \quad a = \{a_{mn}\}_{m,n}$$

with one or more indices. The continuous variable case corresponds to *analogue* signals, while the discrete case corresponds to *digital* signals. Whether continuous or discrete, the single variable should usually be thought of as *time or space*; on the other hand, two variables should always be thought of as *space* variables so that the function or signal then corresponds to an *image*, a *still* image that is, since moving images will involve three or more variables.

(1.1) Energy. The ‘size’ of a function can be measured in many ways - the absolute maximum or absolute minimum value was a standard way in calculus, for instance. An *average* rather than a *pointwise* measure will be more useful to us, however.

As always we’ll take as our starting point the N -dimensional space \mathbb{C}^N of all N -tuples $z = (z_1, z_2, \dots, z_N)$ of complex numbers where ‘size’ or ‘Energy’ of z means the square

$$E(z) = |z_1|^2 + |z_2|^2 + \dots + |z_N|^2 = (\|z\|_{\mathbb{C}^N})^2$$

of the usual Euclidean length. A point of view more rooted in algebraic structures, however, points the way to the future. Let’s denote by \mathbb{Z}_N the set

$$\mathbb{Z}_N = \{0, 1, 2, \dots, N-1\}$$

of integers; you may have learned already in algebraic structures that this set has an additive structure defined on it, namely addition $m + n = k$ where $m + n \equiv k \pmod{N}$ which in number theory courses you probably called modular addition. A sequence z in \mathbb{C}^N is then just a function $z : \mathbb{Z}_N \rightarrow \mathbb{C}$. But to appreciate why modular addition arises, set $w = e^{2\pi i/N}$ so that $w^N = 1$; then $w^m w^n = w^{m+n} = w^k$. This will become fundamental when we want to apply Fourier analysis to finite data sets.

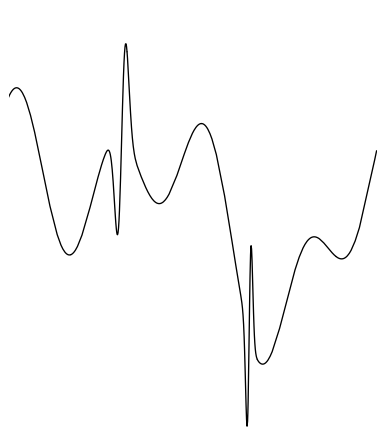
There are natural extensions to functions $f : X \rightarrow \mathbb{C}$ defined on any set X on which some notion of integration or summation and are defined. For example, when $X = \mathbb{R} = (-\infty, \infty)$ or $X = \mathbb{R}^2$ the *Energy* of an analogue signal f is defined by

$$E(f) = \int_{-\infty}^{\infty} |f(x)|^2 dx, \quad \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |f(x, y)|^2 dx dy.$$

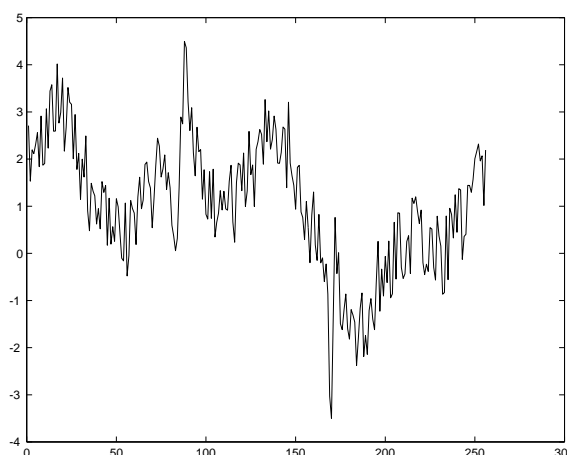
while in the case of a (two-way) infinite digital signal a defined on $X = \mathbb{Z}$ or on \mathbb{Z}^2 the energy is given by

$$E(a) = \sum_{n=-\infty}^{\infty} |a_n|^2, \quad \sum_{m,n=-\infty}^{\infty} |a_{mn}|^2$$

respectively; addition on \mathbb{R} , \mathbb{R}^2 or on \mathbb{Z} , \mathbb{Z}^2 is well-known! A *finite energy* function (or signal) is thus one for which $E(f) < \infty$. In the analogue case, the set of all finite energy functions $f = X \rightarrow \mathbb{C}$ will be denoted by $L^2(X)$, and by $\ell^2(X)$ in the discrete case of digital signals. In particular, therefore, we can, and often do, write $\ell^2(\mathbb{Z}_N)$ instead of \mathbb{C}^N . Typical examples of such functions/signals are

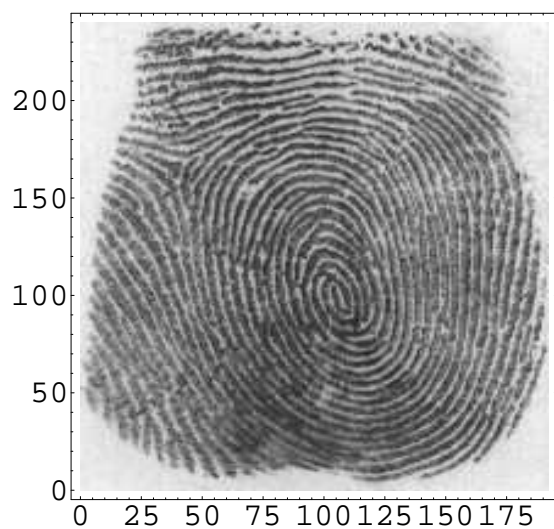


‘transients’

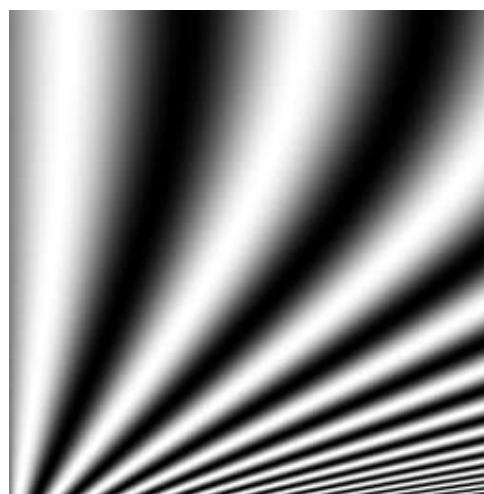


‘transients+noise’

in the one variable case, and



‘real data’



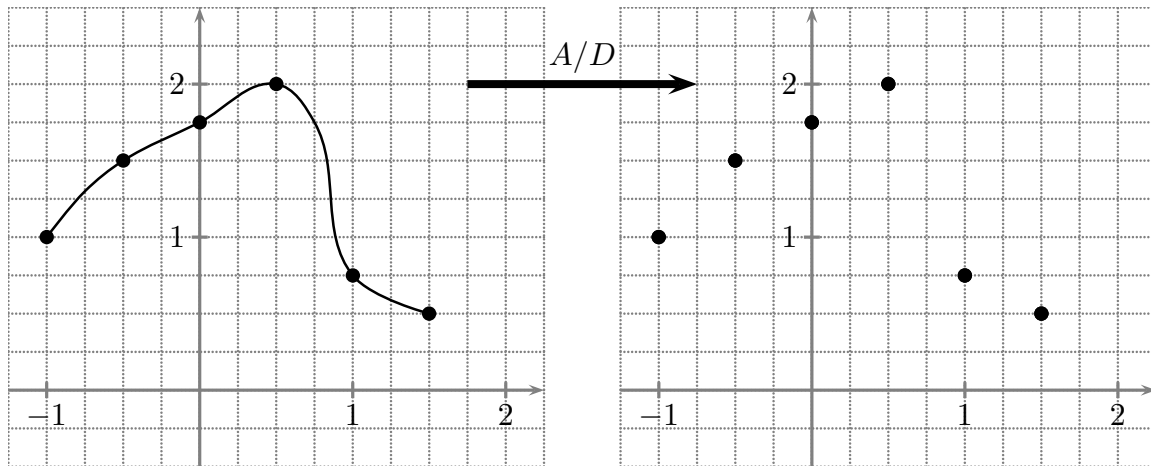
‘synthetic data’

in the two variable case.

(1.2) Analogue to digital. Many discrete signals come from *sampling* continuous functions:

$$f(t) \longrightarrow \{a_n\}_n, \quad a_n = f\left(\frac{n}{2^k}\right)$$

where the integer n varies but the integer k is fixed. Here we have sampled f by evaluating the function *uniformly at dyadic rationals*, the fixed choice of k indicating the *resolution*, the larger k is the *finer* the resolution.



Such *analogue to digital* (A/D) conversion is central to communications technology. But does A/D conversion *necessarily* lead to a loss in energy?

(1.3) Digital to analogue. At the heart of wavelet analysis is the reversal of A/D conversion. Fix a function $\phi = \phi(x)$ and for each pair of integers k, n define a new function

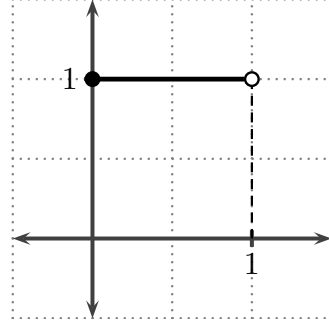
$$\phi_{kn}(x) = 2^{k/2} \phi(2^k x - n).$$

The inclusion of the $2^{k/2}$ -factor preserves energy since

$$E(\phi_{kn}) = 2^k \int_{-\infty}^{\infty} |\phi(2^k x - n)|^2 dx = \int_{-\infty}^{\infty} |\phi(y)|^2 dy = E(\phi)$$

after a change of variable $x \rightarrow y = 2^k x - n$ (notice how we are using structural such as addition and scaling properties of the set X on which ϕ is defined. A simple example and a few graphs suggest what the significance of ϕ_{kn} is. Consider the *box* function

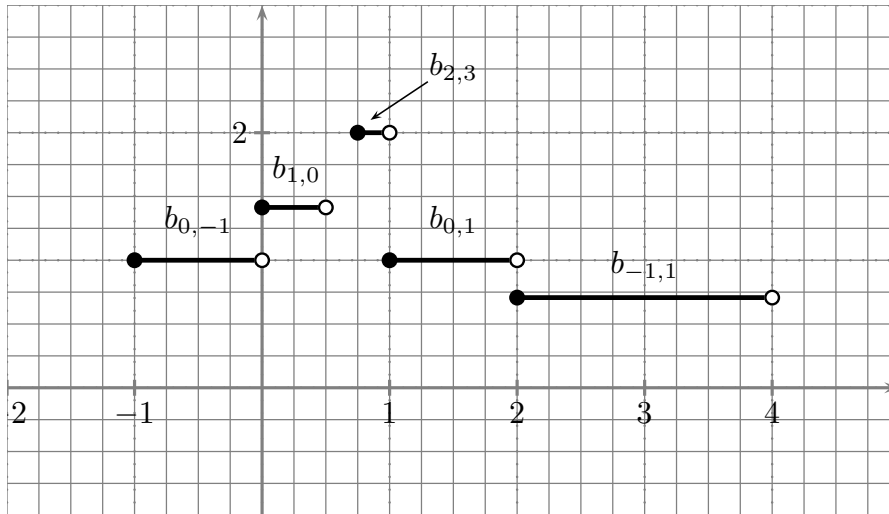
$$b(x) = \begin{cases} 1, & 0 \leq x < 1, \\ 0, & \text{elsewhere.} \end{cases}$$



Then $b(x)$ lives on $[0, 1)$, whereas $b(2^k x)$ lives on $[0, 1/2^k)$, *i.e.*, $b(2^k x)$ is a ‘speeded up’ or ‘slowed down’ version of $b(x)$ according as $k > 0$ or $k < 0$; on the other hand, $b(x - n)$ lives on $[n, n + 1)$, *i.e.*, $b(x - n)$ is a ‘delayed’ or ‘advanced’ version of $b(x)$ according as $n > 0$ or $n < 0$. Putting the two together we see that $b_{kn}(x)$ lives on $[2^{-k}n, 2^{-k}(n + 1))$; in general mathematical terms,

$$\text{supp}(b_{kn}) = \left[\frac{n}{2^k}, \frac{n+1}{2^k} \right).$$

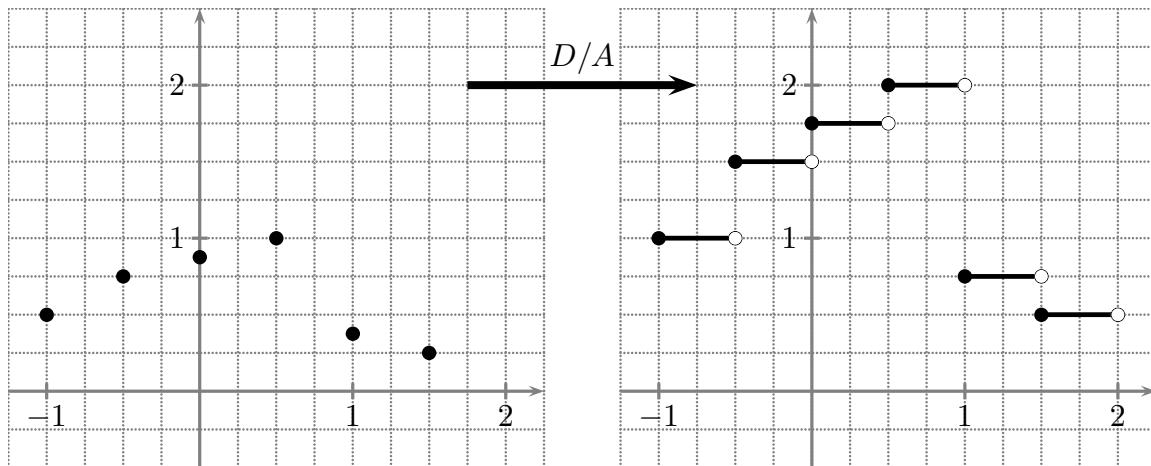
For fixed k , therefore, we get a *specific resolution*, while varying n *changes location*. Some examples of the b_{kn} are shown in



Given a sequence $a = \{a_n\}$ we can now associate a function

$$f(x) = \sum_n a_n b_{kn}(x)$$

at each fixed resolution k . For example at resolution $k = 1$, we get



In addition,

$$E(f) = \int_{-\infty}^{\infty} \left| \sum_n a_n b_{kn}(x) \right|^2 dx = \sum_n |a_n|^2 \int_{-\infty}^{\infty} |b_{kn}(x)|^2 dx = E(a),$$

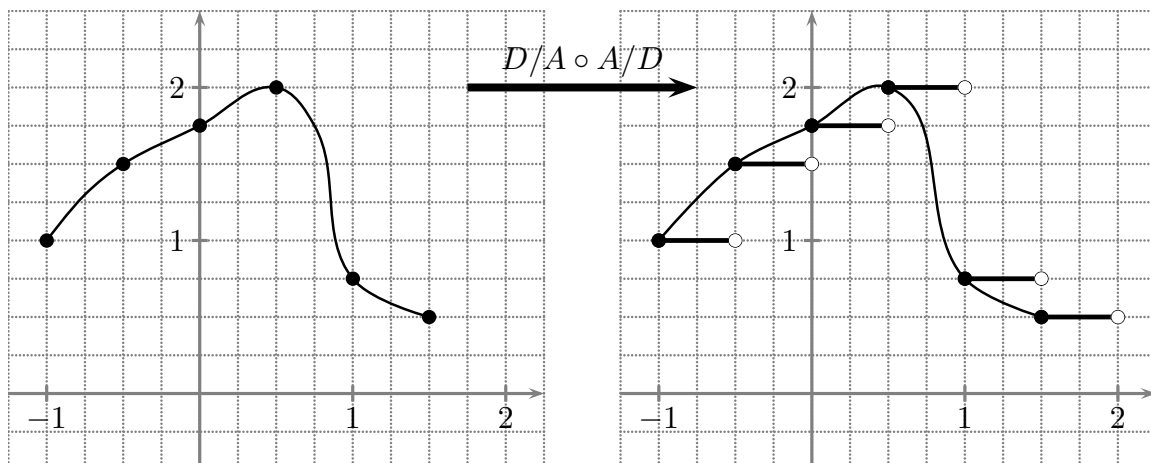
since

$$\int_{-\infty}^{\infty} b_{km}(x) \overline{b_{kn}(x)} dx = \begin{cases} 1, & m = n, \\ 0, & m \neq n, \end{cases}$$

(see problems). Consequently, this D/A conversion $a \rightarrow f$ using the box function has preserved energy. Notice that the composition

$$(D/A) \circ (A/D) : f \rightarrow \sum_n f\left(\frac{n}{2^k}\right) b_{kn}(x)$$

takes a function to an approximating function at a fixed resolution k . As the example, again at $k = 1$ we get



shows, however, using the box function gives a poor visual approximation which may not improve much even at high resolution because of the discontinuity of $b(x)$. This begins to suggest that continuous functions will do a better job for us; perhaps differentiable functions will give an even greater improvement - the eye can perceive slope and concavity!

(1.4) Inner product spaces. Just as functions can be added to form a new function, or vectors in 3-space added to form a new vector, so the sum of two signals makes good sense. Doubling the strength of a signal shows that (real) multiples of a signal also is a naturally occurring concept exactly as it is for functions and vectors. All this suggests that collections of functions, vectors, and signals share two common features: they have length (energy) and linear combinations of them can be formed. As the mathematical notion of *vector space* captures the idea of linear combinations, what we need to do is add a notion of length to a vector space. With the basic idea from calculus of *dot (or inner) product* in mind, let's introduce *inner product spaces*.

Definition. *let \mathcal{V} be a (real or complex) vector space. An inner product on \mathcal{V} is a mapping $(\cdot, \cdot) : \mathcal{V} \times \mathcal{V} \rightarrow \mathbb{C}$ such that*

- (i) $(f, f) \geq 0$, $(f, f) = 0$ if and only if $f = 0$, $(f, g) = \overline{(g, f)}$,
- (ii) $(f + g, h) = (f, h) + (g, h)$, $(\lambda f, g) = \lambda(f, g)$,

for all f, g, h in \mathcal{V} and λ in \mathbb{C} . The Energy of f in \mathcal{V} is defined by $E(f) = (f, f)$. When \mathcal{V} is a real vector space we require that $(\cdot, \cdot) : \mathcal{V} \times \mathcal{V} \rightarrow (-\infty, \infty)$ and that $(f, g) = (g, f)$.

The usual *dot product*

$$\mathbf{u} \cdot \mathbf{v} = (u_1, u_2, u_3) \cdot (v_1, v_2, v_3) = u_1 v_1 + u_2 v_2 + u_3 v_3$$

on vectors makes 3-space into a (real) inner product space. The energy

$$E(\mathbf{u}) = u_1^2 + u_2^2 + u_3^2 = (\mathbf{u}, \mathbf{u})$$

of \mathbf{u} is then simply the square of the usual notion of length, $\|\mathbf{u}\|$, of \mathbf{u} , agreeing with the earlier definition of energy for a signal $\{u_1, u_2, u_3\}$. In the same way, the spaces of finite energy signals introduced earlier become inner product spaces under the respective inner products

- (i) $\ell^2(\mathbb{Z}_N) = \mathbb{C}^N$: for sequences $z = \{z_n\}_n$, $w = \{w_n\}_n$

$$(z, w) = \sum_{n=0}^{N-1} z_n \overline{w_n};$$

(ii) ℓ^2 : for sequences $a = \{a_n\}_n$, $b = \{b_n\}_n$

$$(a, b) = \sum_{n=-\infty}^{\infty} a_n \overline{b_n};$$

(iii) $\ell^2(\mathbb{Z}^2)$: for sequences $a = \{a_{mn}\}_{mn}$, $b = \{b_{mn}\}_{mn}$

$$(a, b) = \sum_{m, n=-\infty}^{\infty} a_{mn} \overline{b_{mn}};$$

(iv) $L^2[0, 1]$: for functions $f = f(x)$, $g = g(x)$

$$(f, g) = \int_0^1 f(x) \overline{g(x)} dx,$$

and correspondingly for functions on any interval $[a, b]$.

(v) $L^2(-\infty, \infty)$: for functions $f = f(x)$, $g = g(x)$

$$(f, g) = \int_{-\infty}^{\infty} f(x) \overline{g(x)} dx,$$

and correspondingly for functions of several variables.

It will be convenient to collect together some properties of an inner product.

Theorem. *Let \mathcal{V} be a complex inner product space. Then*

(i) (*Cauchy-Schwarz inequality*)

$$|(f, g)|^2 \leq E(f) E(g),$$

(ii) (*Triangle inequality*)

$$E^{1/2}(f + g) \leq E^{1/2}(f) + E^{1/2}(g),$$

(iii) (*Polarization identity*)

$$(f, g) = \frac{1}{4} \left\{ E(f + g) - E(f - g) + iE(f + ig) - iE(f - ig) \right\}$$

hold for all f, g in \mathcal{V} .

Exactly the same results hold for a real inner product space except that the polarization identity simplifies to

$$(f, g) = \frac{1}{4} \{ E(f+g) - E(f-g) \}.$$

Proof of Theorem: (i) When $g = 0$ the inequality is obvious, so we can assume that $g \neq 0$. Thus it is enough to show that

$$\left| \left(f, \frac{g}{E^{1/2}(g)} \right) \right|^2 \leq E(f),$$

or, equivalently, that

$$E(g) = 1 \implies |(f, g)|^2 \leq E(f).$$

Now when $E(g) = 1$,

$$\begin{aligned} 0 &\leq E(f - (f, g)g) = (f - (f, g)g, f - (f, g)g) \\ &= E(f) + |(f, g)|^2 - (f, (f, g)g) - ((f, g)g, f). \end{aligned}$$

But

$$(f, (f, g)g) + ((f, g)g, f) = 2|(f, g)|^2.$$

Consequently,

$$E(f) - |(f, g)|^2 \geq 0, \quad E(g) = 1,$$

establishing the Cauchy-Schwarz inequality.

(ii) For any f, g in \mathcal{V} ,

$$\begin{aligned} 0 &\leq E(f+g) = (f+g, f+g) = E(f) + E(g) + (f, g) + (g, f) \\ &= E(f) + E(g) + (f, g) + \overline{(f, g)}. \end{aligned}$$

Consequently, in view of the Cauchy-Schwarz inequality,

$$0 \leq E(f+g) \leq E(f) + E(g) + 2|(f, g)| \leq (E^{1/2}(f) + E^{1/2}(g))^2,$$

establishing also the triangle inequality.

(iii) Left as exercise (see problems) \square

(1.5) Orthonormal families. Just as the dot product on 3-vectors has an equivalent geometric formulation

$$\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\| \|\mathbf{v}\| \cos \theta,$$

with θ the angle between \mathbf{u} and \mathbf{v} , the inner product (f, g) provides also a measure of the ‘angle’ between f and g for a general inner product space. From an analytic point of view, the inner product (f, g) of two functions provides us with a measure of how well f *correlates* with g because if f looks like g then (f, g) will be close to the energy $E(g)$ of g ; for instance, if g oscillates, then f will correlate well with g when f is oscillating like g .

The geometric interpretation of the dot product allowed us to introduce *perpendicular* unit vectors

$$\mathbf{i} = (1, 0, 0), \quad \mathbf{j} = (0, 1, 0), \quad \mathbf{k} = (0, 0, 1)$$

in 3-space, (aka *orthogonal*), so we can do the same in a general \mathcal{V} . This is another of the really key ideas in wavelets.

Definition. A subset $\{\phi_n\}$ of an inner product space \mathcal{V} is said to be an *orthonormal family* when

$$E(\phi_n) = 1, \quad (\phi_m, \phi_n) = 0 \quad m \neq n,$$

i.e., the family consists of mutually orthogonal elements having unit energy.

For example, when the sequence $\varepsilon^{(n)} = \{\varepsilon_\ell^{(n)}\}_\ell$ is defined by

$$\varepsilon_\ell^{(n)} = \begin{cases} 1, & \ell = n, \\ 0, & \ell \neq n, \end{cases}$$

the family $\{\varepsilon^{(n)} : n = \dots, -1, 0, 1, \dots\}$ is orthonormal in ℓ^2 ; it is an obvious generalization of the orthonormal family $\mathbf{i}, \mathbf{j}, \mathbf{k}$ for 3-space. But a vector \mathbf{v} in 3-space can be represented as

$$\mathbf{v} = (v_1, v_2, v_3) = v_1\mathbf{i} + v_2\mathbf{j} + v_3\mathbf{k}$$

where

$$v_1 = (\mathbf{v}, \mathbf{i}), \quad v_2 = (\mathbf{v}, \mathbf{j}), \quad v_3 = (\mathbf{v}, \mathbf{k});$$

naturally, there is a corresponding representation for any sequence $a = \{a_n\}$ in the infinite-dimensional case ℓ^2 :

$$a = \sum_n a_n \varepsilon^{(n)} = \sum_n (a, \varepsilon^{(n)}) \varepsilon^{(n)}.$$

This use of an orthonormal family to represent a signal in ℓ^2 carries over to an arbitrary inner product space.

Definition. For a function f in an inner product space \mathcal{V} the **orthonormal series**, $S[f]$, of f associated with an orthonormal family $\{\phi_n\}$ in \mathcal{V} is defined by

$$S[f] = \sum_n (f, \phi_n) \phi_n.$$

The numerical values (f, ϕ_n) are called the **coefficients** of f .

Notice that we don't claim that $f = S[f]$ automatically holds. For instance, the family $\{\mathbf{i}, \mathbf{j}\}$ is orthonormal in 3-space, but no vector in the z -direction can be represented in terms of $\{\mathbf{i}, \mathbf{j}\}$ since these vectors lie in the x - y plane; we obviously don't have a large enough family of orthonormal vectors! Making precise when $f = S[f]$ is a key step in the theory.

Bessel's Inequality. Let $\{\phi_n\}$ be an orthonormal family in an inner product space \mathcal{V} . Then the inequality

$$\sum_n |(f, \phi_n)|^2 \leq E(f)$$

holds for all f in \mathcal{V} ; in other words, the sequence $\{(f, \phi_n)\}_n$ of coefficients has finite energy as a sequence in ℓ^2 , and $E(\{(f, \phi_n)\}_n) \leq E(f)$.

Proof: using the symmetry and linearity properties of the inner product we see that

$$\begin{aligned} 0 &\leq E\left(f - \sum_{n=0}^N (f, \phi_n) \phi_n\right) = \left(f - \sum_{m=0}^N (f, \phi_m) \phi_m, f - \sum_{n=0}^N (f, \phi_n) \phi_n\right) \\ &= E(f) - 2 \sum_{n=0}^N |(f, \phi_n)|^2 + \sum_{m,n=0}^N (f, \phi_m) \overline{(f, \phi_n)} (\phi_m, \phi_n). \end{aligned}$$

But by orthonormality,

$$\sum_{m,n=0}^N (f, \phi_m) \overline{(f, \phi_n)} (\phi_m, \phi_n) = \sum_{n=0}^N |(f, \phi_n)|^2.$$

Consequently,

$$E(f) - \sum_{n=0}^N |(f, \phi_n)|^2 \geq 0$$

for all choices of N . From this Bessel's inequality follows immediately letting $N \rightarrow \infty$, making obvious changes if the orthonormal family is indexed differently. \square

In the example of $\{\mathbf{i}, \mathbf{j}\}$ in 3-space, Bessel's inequality becomes a *strict* inequality

$$|(\mathbf{v}, \mathbf{i})|^2 + |(\mathbf{v}, \mathbf{j})|^2 < E(\mathbf{v})$$

when \mathbf{v} has a component in the z -direction, *i.e.*, in the \mathbf{k} -direction. This suggests what we have to do to guarantee that $f = S[f]$.

Definition. An orthonormal family $\{\phi_n\}$ in \mathcal{V} is said to be **complete** when the equality

$$\sum_n |(f, \phi_n)|^2 = E(f)$$

holds for all f in \mathcal{V} ; in other words, when Bessel's inequality becomes an equality for all f .

Given a complete orthonormal family $\{\phi_n\}_{n=1}^{\infty}$ it can be shown that

$$S[f] = \sum_n (f, \phi_n) \phi_n = f$$

in the sense

$$(\dagger) \quad \lim_{N \rightarrow \infty} E\left(f - \sum_{n=1}^N (f, \phi_n) \phi_n\right) = 0.$$

When the orthonormal family is indexed by something other than $\{1, 2, \dots\}$, the partial sums have to be changed, but it's usually clear how to do that. For instance, if the family is indexed by $\{\dots, -1, 0, 1, \dots\}$, all the way from $-\infty$ to ∞ , then we would replace (\dagger) with

$$(\ddagger) \quad \lim_{N \rightarrow \infty} E\left(f - \sum_{n=-N}^N (f, \phi_n) \phi_n\right) = 0.$$

Given the importance of these ideas, let's collect them together in one result.

Theorem. If $\{\phi_n\}_n$ is a complete orthonormal family in an inner product space \mathcal{V} , then each f in \mathcal{V} admits an orthonormal series representation

$$f = S[f] = \sum_n (f, \phi_n) \phi_n.$$

Furthermore, the coefficients $\{(f, \phi_n)\}_n$ satisfy the identities

(i) (Plancherel)

$$E(f) = \sum_n |(f, \phi_n)|^2,$$

(ii) (Parseval)

$$(f, g) = \sum_n (f, \phi_n) \overline{(g, \phi_n)}$$

for all f, g in \mathcal{V} .

Parseval's equation follows immediately from Plancherel's theorem by polarization (see proof of a more general result in (1.7)), while Plancherel's theorem is just the same as our definition of completeness!

(1.6) Examples. 1. The simplest generalization of the $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ basis of \mathbb{R}^3 is the so-called *Standard Basis*

$\varepsilon^{(1)} = (1, 0, 0, \dots, 0, 0)$, $\varepsilon^{(2)} = (0, 1, 0, \dots, 0, 0)$, \dots , $\varepsilon^{(N)} = (0, 0, 0, \dots, 0, 1)$ of $\ell^2(\mathbb{Z}_N)$ and its generalization to the complete orthonormal family

$$\{\varepsilon^{(n)} : -\infty < n < \infty\}$$

in ℓ^2 we have met already. This family will also be referred as the *Standard Basis*.

2. Of crucial importance to signal analysis is the so-called *Fourier Basis*: set $w = e^{2\pi i/N}$. Then each power w^k , $k \geq 0$, is an N^{th} root of unity in the sense that $w^N = 1$. Now set

$$\xi^{(0)} = \frac{1}{\sqrt{N}}(1, 1, 1, \dots, 1), \quad \xi^{(1)} = \frac{1}{\sqrt{N}}(1, w, w^2, \dots, w^{N-1}),$$

\vdots

$$\xi^{(k)} = \frac{1}{\sqrt{N}}(1, w^k, w^{2k}, \dots, w^{k(N-1)}),$$

\vdots

$$\xi^{(N-1)} = \frac{1}{\sqrt{N}}(1, w^{N-1}, w^{2(N-1)}, \dots, w^{(N-1)^2})$$

define N elements of $\ell^2(\mathbb{Z}_N)$ such that $E(\xi^{(k)}) = 1$ for each k . On other hand, by using the geometric series fact:

$$1 + w^n + w^{2n} + \dots + w^{n(N-1)} = \frac{1 - w^{nN}}{1 - w} = 0,$$

it can be shown that the family

$$\xi^{(0)}, \quad \xi^{(1)}, \quad \xi^{(2)}, \quad \dots, \quad \xi^{(N-1)}$$

is a complete orthonormal family in $\ell^2(\mathbb{Z}_N)$.

3. In \mathbb{R}^2 the vectors

$$\frac{1}{\sqrt{2}}(\mathbf{i} + \mathbf{j}), \quad \frac{1}{\sqrt{2}}(\mathbf{i} - \mathbf{j})$$

form an orthonormal basis because they are just the rotation through 45° of the standard $\{\mathbf{i}, \mathbf{j}\}$ -basis in the plane. This two dimensional example generalizes easily to ℓ^2 in the same way that $\{\varepsilon^{(n)}\}_n$ generalizes $\{\mathbf{i}, \mathbf{j}\}$. Indeed, define $\varphi^{(n)}$, $\tilde{\varphi}^{(n)}$ in ℓ^2 by

$$\varphi^{(n)} = \frac{1}{\sqrt{2}}(\varepsilon^{(2n)} + \varepsilon^{(2n+1)}), \quad \tilde{\varphi}^{(n)} = \frac{1}{\sqrt{2}}(\varepsilon^{(2n)} - \varepsilon^{(2n+1)}).$$

Then each of $\{\varphi^{(n)}\}_n$ and $\{\tilde{\varphi}^{(n)}\}_n$ is orthonormal in ℓ^2 ; in addition, together the family $\{\varphi^{(n)}, \tilde{\varphi}^{(n)}\}_n$ is a complete orthonormal family in ℓ^2 . Consequently, every x in ℓ^2 admits a representation

$$x = \sum_n (x, \varphi^{(n)}) \varphi^{(n)} + \sum_n (x, \tilde{\varphi}^{(n)}) \tilde{\varphi}^{(n)}.$$

In the signal processing literature, such bases were called block bases (can you see why?), and were very common prior to the appearance of wavelets. We shall meet these families $\{\varphi^{(n)}\}_n$, $\{\tilde{\varphi}^{(n)}\}_n$ later in lecture 2 and crucially in lecture 4 in connection with filter banks.

4. Since

$$\int_a^{a+1} e^{2\pi im\xi} e^{-2\pi in\xi} d\xi = \begin{cases} 1, & m = n, \\ 0, & m \neq n, \end{cases}$$

the family $\{e^{2\pi in\xi} : -\infty < n < \infty\}$ of period 1 functions is orthonormal in $L^2[a, a+1]$ for any choice of a ; in particular, it is orthonormal in $L^2[-\frac{1}{2}, \frac{1}{2}]$ as well as in $L^2[0, 1]$. It can be shown that this family is also complete in $L^2[a, a+1]$ for any choice of a .

5. The family

$$\{b_{0n} : -\infty < n < \infty\}, \quad b_{0n}(x) = b(x-n),$$

of all integer delays of the box function b is orthonormal in $L^2(-\infty, \infty)$ because

$$\int_{-\infty}^{\infty} b_{0m}(x) \overline{b_{0n}(x)} dx = \int_m^{m+1} b(x-n) dx = \begin{cases} 1, & m = n, \\ 0, & m \neq n. \end{cases}$$

As any combination $\sum_n a_n b(x-n)$ is constant on intervals of length 1, however, there is no way that the Haar function, for instance, can be written as $\sum_n a_n b(x-n)$. Hence $\{b_{0n} : -\infty < n < \infty\}$ is not complete in $L^2(-\infty, \infty)$. Similarly, the family

$$\{b_{kn} : -\infty < n < \infty\}, \quad b_{kn}(x) = 2^{k/2} b(2^k x - n),$$

is orthonormal, but not complete, in $L^2(-\infty, \infty)$, for each fixed resolution k .

6. The family

$$\{h_{0n} : -\infty < n < \infty\}, \quad h_{0n}(x) = h(x-n),$$

of all integer delays of the Haar function h is orthonormal in $L^2(-\infty, \infty)$. Since the integer translate $h_{0n} = h(x-n)$ lives on $[n, n+1)$ and has average 0 on that interval, the

box function b cannot be written as a linear combination $\sum_n a_n b(x - n)$; in particular, therefore, the family $\{h_{0n}\}$ is not complete in $L^2(-\infty, \infty)$. Similarly, the family

$$\{h_{kn} : -\infty < n < \infty\}, \quad h_{kn}(x) = 2^{k/2} h(2^k x - n),$$

is orthonormal, but not complete, in $L^2(-\infty, \infty)$, for each fixed resolution k .

7. By allowing all possible resolutions and integer delays of the Haar function, however, we do get a complete orthonormal family in $L^2(-\infty, \infty)$. To be precise, when h_{kn} is defined as in the previous example, then the family $\{h_{kn} : -\infty < k, n < \infty\}$ is orthonormal since calculations show that

$$(h_{jm}, h_{k\ell}) = 2^{(j+k)/2} \int_{-\infty}^{\infty} h(2^j x - m) h(2^k x - \ell) dx = \begin{cases} 1, & j = k, m = \ell, \\ 0, & \text{otherwise;} \end{cases}$$

in addition, the family is complete in $L^2(-\infty, \infty)$ so each finite energy function f admits a representation

$$f(x) = \sum_{k, \ell = -\infty}^{\infty} (f, h_{k\ell}) h_{k\ell}(x).$$

This Haar decomposition is the prototype for all wavelet representations of finite energy analogue signals for reasons we shall make clearer in lecture 4.

(1.7) Energy-preserving mappings, adjoints. In linear algebra a key role is played by $m \times n$ matrices $A = [A_{mn}]$ because they define linear transformations

$$A : x \longrightarrow Ax = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

by (right) matrix multiplication from \mathbb{R}^n into \mathbb{R}^m . The same is true in the complex case, of course. This idea extends to arbitrary linear transformations $T : V \longrightarrow W$ from one vector space V into a possibly different vector space W . When V and W are inner product spaces, however, we shall want such T to ‘respect’ the energy of vectors in V . Let’s make this precise because it too will be important in our development of the theory of wavelets.

Definition. A linear operator $\mathcal{T} : \mathcal{V} \longrightarrow \mathcal{W}$ from one inner product space \mathcal{V} into a possibly different inner product space \mathcal{W} (aka linear mapping) is said to be **bounded** when there is a constant such that the inequality

$$E(\mathcal{T}f) \leq \text{const. } E(f)$$

holds for all f in \mathcal{V} . It is said to be **energy-preserving** when $E(\mathcal{T}f) = E(f)$ for all f .

If $\{\phi_n\}$ is orthonormal in \mathcal{V} , then by Bessel's inequality the coefficient mapping

$$\mathcal{T} : f \longrightarrow \{(f, \phi_n)\}_n$$

is bounded (with constant at most one) from \mathcal{V} into ℓ^2 ; in addition, this coefficient mapping is energy-preserving when $\{\phi_n\}$ is a complete orthonormal family. In the reverse direction, consider the D/A mapping

$$\mathcal{S} : a = \{a_n\}_n \longrightarrow \sum_n a_n b(x - n)$$

from ℓ^2 into $L^2(-\infty, \infty)$. Because the integer translates $b(x - n)$ are orthonormal in $L^2(-\infty, \infty)$, we see that

$$E(\mathcal{S}a) = \int_{-\infty}^{\infty} \left| \sum_n a_n b(x - n) \right|^2 dx = \sum_n |a_n|^2 = E(a).$$

Consequently, \mathcal{S} is energy-preserving as a mapping from finite energy digital signals to finite energy analogue signals. Similarly, the D/A mapping

$$\mathcal{S}_k : a = \{a_n\}_n \longrightarrow \sum_n a_n b_{kn}(x), \quad b_{kn}(x) = 2^{k/2} b(2^k x - n),$$

is energy-preserving at every fixed resolution k because the family $\{b_{kn}\}_n$ is orthonormal for each fixed k .

The previous examples show the close relationship between energy-preserving operators and orthonormal families. We need to make this precise in general. Suppose then that $\mathcal{T} : \mathcal{V} \longrightarrow \mathcal{W}$ is an energy-preserving operator from one inner product space into a possibly different one \mathcal{W} . By the polarization identity, \mathcal{T} preserves inner products also since

$$\begin{aligned} (\mathcal{T}f, \mathcal{T}g) &= \frac{1}{4} \left\{ E(\mathcal{T}(f + g)) - E(\mathcal{T}(f - g)) + iE(\mathcal{T}(f + ig)) - iE(\mathcal{T}(f - ig)) \right\} \\ &= \frac{1}{4} \left\{ E(f + g) - E(f - g) + iE(f + ig) - iE(f - ig) \right\} = (f, g) \end{aligned}$$

(linearity of \mathcal{T} ensures that $\mathcal{T}f + \mathcal{T}g = \mathcal{T}(f + g)$ etc). Hence, if $\{\phi_n\}_n$ is orthonormal in \mathcal{V} ,

$$(\mathcal{T}\phi_m, \mathcal{T}\phi_n) = (\phi_m, \phi_n) = \begin{cases} 1, & m = n, \\ 0, & m \neq n, \end{cases}$$

proving the following result.

Theorem. An energy-preserving mapping $\mathcal{T} : \mathcal{V} \longrightarrow \mathcal{W}$ from one inner product space into a possibly different one \mathcal{W} maps an orthonormal family $\{\phi_n\}$ in \mathcal{V} to an orthonormal family $\{\mathcal{T}(\phi_n)\}_n$ in \mathcal{W} .

Finally, recall the idea of the *adjoint*

$$A^* = \overline{A}^t = \begin{bmatrix} \overline{a_{11}} & \overline{a_{21}} & \dots & \overline{a_{m1}} \\ \overline{a_{12}} & \overline{a_{22}} & \dots & \overline{a_{m2}} \\ \vdots & \vdots & & \vdots \\ \overline{a_{1n}} & \overline{a_{2n}} & \dots & \overline{a_{mn}} \end{bmatrix}$$

of the $m \times n$ matrix dealt with earlier. Then A^* is an $n \times m$ matrix, hence it defines a linear transformation $\mathbb{R}^m \longrightarrow \mathbb{R}^n$ by right matrix multiplication. But is there a deeper meaning to this? The answer lies in the notion of *adjoint* of a bounded operator between inner product spaces.

Definition. Let $T : \mathcal{V} \longrightarrow \mathcal{W}$ be a bounded linear operator from one inner product space into a possibly different one \mathcal{W} . Then the **adjoint** T^* of T is the bounded linear operator $T^* : \mathcal{W} \longrightarrow \mathcal{V}$ such that $(Tf, g) = (f, T^*g)$ holds for all f in \mathcal{V} and g in \mathcal{W} .

When T is determined by an $m \times n$ matrix A , the adjoint of T is determined by the adjoint matrix A^* of A (see problems).

A particularly interesting example for us is the adjoint of the *coefficient* mapping

$$T : \mathcal{V} \longrightarrow \ell^2, \quad Tf = \{(f, \phi_n)\}_n$$

determined by an orthonormal family $\{\phi_n\}_n$ in \mathcal{V} . For then

$$T^* : \{a_n\}_n \longrightarrow \sum_n a_n \phi_n$$

since

$$(Tf, a) = \sum_n (f, \phi_n) \overline{a_n} = \left(f, \sum_n a_n \phi_n \right) = (f, T^*a).$$

Notice that T^* is always energy-preserving whether or not $\{\phi_n\}_n$ is *complete* in \mathcal{V} . We have used this several times already. For instance, the family $\{b_{0n}\}_n$, $b_{0n}(x) = b(x - n)$, is orthonormal in $L^2(-\infty, \infty)$, and the adjoint of the coefficient mapping

$$f \longrightarrow \{(f, b_{0n})\}_n, \quad (f, b_{0n}) = \int_n^{n+1} f(y) dy$$

is simply the *D/A* mapping

$$\mathcal{S} : \{a_n\}_n \longrightarrow \sum_n a_n b(x - n).$$

(1.8) Summary, compression. So what have we achieved so far? It has been convenient to think of collections of signals, whether analogue or digital, and functions, of continuous or discrete variables, as belonging to an inner product space. This combined the linear structure of signals with the notion of energy and correlation. Families $\{\phi_n\}_n$ of particular elements of that space, orthonormal with respect to the inner product, could then be chosen with the crucial idea of *representing* a signal as an orthonormal series $f = \sum_n (f, \phi_n) \phi_n$ when the family is complete. Why bother? Well, it will surely be helpful if the members of that orthonormal family somehow *capture the fundamental properties of the model* we are studying - they are to be the basic building blocks for the model. Furthermore, completeness ensures that the energy of f is completely determined by the energy of the coefficients $\{(f, \phi_n)\}_n$. There are then two things we can do using these coefficients:

- (i) by dropping *very small* coefficients from $\sum_n (f, \phi_n) \phi_n$ we get only an approximation to f , but the terms $(f, \phi_n) \phi_n$ omitted will have *very small energy*, so presumably the approximation may still be a ‘good’ representation of f . In other words, we have represented the original f , up to a very small error, by fewer terms, thus providing *compression* of f .
- (ii) If the orthonormal family captures key features of a model - say, sharp spikes or sudden oscillations - then the large coefficients $\{(f, \phi_n)\}_n$ in the representation $f = \sum_n (f, \phi_n) \phi_n$ probably identify sharp edges in a signal, thus providing *feature detection*: think ‘little wave’ property.

The four signals pictured in section (1.1) illustrate these ideas well. For example, detection of transients could be an important issue for the first of the one-dimensional signals, while for the two fingerprint pictures the question might be one of detecting or characterizing edges. For efficient transmission or storage of fingerprints, however, compression would be crucial. In practice, say in medical imaging, most transients or edges, however, would be hidden in lots of noise as in the second of the one-dimensional signals. So the question would become one of isolating features within noise; or one could simply ask for de-noising. Whatever the question, however, the ‘moral of the story’ is always the same: everything depends on the choice of orthonormal family $\{\phi_n\}_n$.

Problems.

1. Establish the parallelogram law

$$E(f + g) + E(f - g) = 2E(f) + 2E(g), \quad f, g \in \mathcal{V}$$

for an inner product space \mathcal{V} . Why is this called the parallelogram law?

2. Establish the polarization identity

$$(f, g) = \frac{1}{4} \left\{ E(f + g) - E(f - g) + iE(f + ig) - iE(f - ig) \right\}$$

for f, g in an inner product space.

There are many examples of inner product spaces. The next three problems discuss examples quite different from the ones given in this lecture.

3. Recall that the trace of a 2×2 matrix is defined by

$$\text{trace}(A) = a + d, \quad A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

Use this to show that the set $\mathbb{C}^{2 \times 2}$ of all such 2×2 matrices having complex entries becomes an inner product space with respect to

$$(A, B) = \text{trace}(AB^*).$$

Exhibit at least one family in $\mathbb{C}^{2 \times 2}$ which is complete and orthonormal with respect to this inner product.

4. Modify the inner product on 2×2 matrices introduced in problem 3 so that the family $\mathbb{C}^{2 \times 3}$ of 2×3 matrices having complex entries becomes an inner product space. (Don't forget to show that $(A, B) = \overline{(B, A)}$ for all 2×3 -matrices A, B .)

5. Denote by \mathcal{A} the family of all (real-valued) functions f that have a Taylor series expansion

$$f(x) = \sum_{n=0}^{\infty} \frac{1}{n!} f^{(n)}(0) x^n$$

whose interval of convergence is some interval about the origin, the interval of convergence being allowed to vary with f . Notice that

$$f(D) = \sum_{n=0}^{\infty} \frac{1}{n!} f^{(n)}(0) D^n, \quad D = \frac{d}{dx},$$

then makes sense as a differential operator. Show that \mathcal{A} becomes an inner product space with respect to the inner product

$$(f, g) = f(D)g(x) \Big|_{x=0};$$

and that the corresponding energy $E(f)$ of a function in \mathcal{A} is given by

$$E(f) = \sum_{n=0}^{\infty} \frac{1}{n!} (f^{(n)}(0))^2.$$

(a) Show that the family $\{\frac{1}{\sqrt{n!}} x^n : n = 0, 1, \dots, \}$ is orthonormal and complete in \mathcal{A} with respect to this inner product.

(b) Interpret the degree N Taylor polynomial

$$P_N(f, x) = \sum_{n=0}^N \frac{1}{n!} f^{(n)}(0) x^n$$

as a ‘compression’ of f . What is the loss in energy by using this compression? Do you think this is a good measure of the compression of f ?

6. Show that every 2×2 matrix A acting by right matrix multiplication

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} : z = (z_1 \ z_2) \longrightarrow (az_1 + bz_2, cz_1 + dz_2)$$

on \mathbb{C}^2 automatically defines a bounded operator on \mathbb{C}^2 such that

$$E(Az) \leq E(A) E(z), \quad x = (z_1, z_2) \in \mathbb{C}^2.$$

(Hint: apply Cauchy-Schwarz to the entries in Az).

The next few problems establish orthonormality properties of the functions

$$b_{kn}(x) = 2^{k/2} b(2^k x - n), \quad h_{kn}(x) = 2^{k/2} h(2^k x - n)$$

obtained by stretching and translating the box and Haar functions. One of the best ways of providing careful proofs of these properties as well as beginning to develop a good understanding of the more general wavelet functions

$$\phi_{kn}(x) = 2^{k/2} \phi(2^k x - n), \quad \psi_{kn}(x) = 2^{k/2} \psi(2^k x - n)$$

defined later is to look at the geometry and number theory associated with dyadic intervals. Recall from the lecture that to each pair of integers k, n there corresponds a dyadic interval

$$I_{kn} = [2^{-k}n, 2^{-k}(n+1)).$$

Often we get ‘clever’ and denote such intervals by I, I' etc. when specific values of k, n aren’t needed.

7. Establish the following results for dyadic intervals.

- (a) Show that the functions b_{kn} and h_{kn} 'live on' I_{kn} in the sense that they are ZERO off I_{kn} .
- (b) Show that the family $\{I_{kn} : -\infty < n, \infty\}$ is a partition of $(-\infty, \infty)$ for each fixed choice of k . (Think of this as the partition of $(-\infty, \infty)$ at *resolution* k .)
- (c) Let $I, I', I \neq I'$ be dyadic intervals such that $I \neq I'$. Show that exactly one of

$$I \subset I', \quad I' \subset I, \quad I \cap I' = \emptyset,$$

holds. Show furthermore that when $I \subset I'$, then either I lies in the first half $[2^{-k}n, 2^{-k}(n + \frac{1}{2})]$ of I' or in the second half $[2^{-k}(n + \frac{1}{2}), 2^{-k}(n + 1)]$ of I' .

8. Show that the family $\{b_{kn} : -\infty < n < \infty\}$ is orthonormal in $L^2(-\infty, \infty)$ in the sense that

$$(b_{km}, b_{kn}) = \begin{cases} 1, & m = n, \\ 0, & m \neq n, \end{cases}$$

for each fixed k .

9. Use property (c) in problem 7 to establish the following results:

- (a) the family $\{h_{kn} : -\infty < k, n < \infty\}$ is orthonormal in $L^2(-\infty, \infty)$ in the sense that

$$(h_{jm}, h_{kn}) = \begin{cases} 1, & j = k, m = n, \\ 0, & \text{otherwise.} \end{cases}$$

- (b) the family $\{h_{jm} : j \geq k, -\infty < m < \infty\}$ is orthogonal to the family $\{b_{kn} : -\infty < n < \infty\}$ for each fixed k in the sense that

$$(h_{jm}, b_{kn}) = 0$$

for all $j \geq k$ and all integers m, n .

10. Let $T : \mathcal{V} \rightarrow \mathcal{W}$ be energy-preserving. Provide an example to show that $\{T(\phi_n)\}_n$ need not be a complete orthonormal family in \mathcal{W} even if $\{\phi_n\}_n$ is a complete orthonormal family in \mathcal{V} .