WAVELETS: Theory and Applications

John E. Gilbert

Lecture Notes

Fall 1997

Version: August 1997

Contents

Part I. Fourier Analysis on Euclidean Space

1. Hilbert space, bases. A Hilbert space is the natural abstraction of very familiar ideas from 3-space:

(1.1)
(i) *the linear structure of vectors,*
(ii) *the inner product (or dot product) of vectors,*
(iii) *the completeness property of real numbers.*

The natural abstraction of (1.1)(i) is a vector space over the scalar field of real numbers $\mathbb{R}$ or complex numbers $\mathbb{C}$ which for us will always be $\mathbb{C}$ unless it is specifically identified as being $\mathbb{R}$. To axiomatize (1.1)(ii), recall that in 3-space the basis-free form of the dot product of vectors $\mathbf{u}, \mathbf{v}$ is defined by

(1.2) $$\mathbf{u}.\mathbf{v} = \|\mathbf{u}\|\,\|\mathbf{v}\|\cos\theta$$

where $\|\mathbf{u}\|, \|\mathbf{v}\|$ denote the length of $\mathbf{u}, \mathbf{v}$, and θ denotes the angle between $\mathbf{u}$ and $\mathbf{v}$. Thus the dot product captures both the idea of length and of angle, and so adds the geometric structure of 3-space to the usual linear structure of vectors. Clearly

(1.3)(i) $$\mathbf{u}.\mathbf{u} \geq 0\ , \qquad \mathbf{u}.\mathbf{u} = 0 \ \text{ if and only if } \ \mathbf{u} = \mathbf{0}\ , \quad \mathbf{u}.\mathbf{v} = \mathbf{v}.\mathbf{u}\ ;$$

the geometric structure is linked to the linear structure through the third property

(1.3)(ii) $$(\mathbf{u}_1 + \mathbf{u}_2).\mathbf{v} = \mathbf{u}_1.\mathbf{v} + \mathbf{u}_2.\mathbf{v}\ , \qquad (\alpha\mathbf{u}).\mathbf{v} = \alpha(\mathbf{u}.\mathbf{v}) \qquad (\alpha \in \mathbb{R})\ .$$

These properties are the ones used to impose an inner product structure on a vector space.

(1.4) Definition. *Let $\mathcal{V}$ be a complex vector space. An inner product on $\mathcal{V}$ is a mapping $(.,.) : \mathcal{V} \times \mathcal{V} \to \mathbb{C}$ such that*

(i) $(f, f) \geq 0$, $\quad (f, f) = 0$ *if and only if* $f = 0$, $\quad (f, g) = \overline{(g, f)}$,

(ii) $(f + g, h) = (f, h) + (g, h)$, $\quad (\alpha f, g) = \alpha(f, g)$

for all f, g, h in $\mathcal{V}$ and α in $\mathbb{C}$. A vector space equipped with such an inner product is called an inner product space or, sometimes, a pre-Hilbert space.

Reversing the procedure used for 3-space we define the length of an element f of an inner product space by

$$\|f\| = (f, f)^{1/2} \ . \tag{1.5}$$

Clearly

$$\|f\| \geq 0 \ , \quad \|\alpha f\| = |\alpha| \, \|f\| \ , \quad \|f\| = 0 \ \text{ if and only if } \ f = 0 \ . \tag{1.6}$$

This notion of length also has all the usual properties of length in 3-space.

(1.7) Theorem. *Let $\mathcal{V}$ be an inner product space. Then*

(i) *(Cauchy-Schwarz inequality)*

$$|(f, g)| \leq \|f\| \, \|g\| \ ,$$

(ii) *(triangle inequality)*

$$\|f + g\| \leq \|f\| + \|g\| \ ,$$

(iii) *(parallelogram law)*

$$\|f + g\|^2 + \|f - g\|^2 = 2(\|f\|^2 + \|g\|^2)$$

(iv) *(polarization identity)*

$$(f, g) = \tfrac{1}{4}\{\|f + g\|^2 - \|f - g\|^2 + i\|f + ig\|^2 - i\|f - ig\|^2\}$$

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

Proof. (i) When $g = 0$ the inequality is obvious, so we can assume $g = 0$. Thus, by (1.6), it is enough to show that

$$\left| \left(f, \frac{g}{\|g\|} \right) \right| \leq \|f\| \ ,$$

or, equivalently, that

$$(*) \qquad |(f,g)| \le \|f\| \qquad (\|g\| = 1) \ .$$

Now, when $\|g\| = 1$,

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

But

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

Consequently

$$\|f\|^2 - |(f,g)|^2 \ge 0 \qquad (\|g\| = 1) \ ,$$

establishing $(*)$.

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

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

Consequently, by the Cauchy-Schwarz inequality,

$$\|f + g\|^2 \le \|f\|^2 + \|g\|^2 + 2|(f,g)| \le (\|f\| + \|g\|)^2 \ ,$$

establishing the triangle inequality.

(iii) Analogous to $(**)$,

$$(***) \qquad \|f - g\|^2 = \|f\|^2 + \|g\|^2 - (f,g) - (g,f) \ .$$

After adding $(**)$ and $(***)$ we obtain the parallelogram law for an inner product space.

(iv) left as exercise. ∎

Properties (1.6) and (1.7)(ii) show that an inner product space is always a *normed space* in the following sense.

(1.8) DEFINITION. *Let $\mathcal{V}$ be a complex vector space. Then $\mathcal{V}$ is said to be a normed space when there is a function $\|.\| : \mathcal{V} \to \mathbb{R}$ such that*

(i) $\|f\| \geq 0$, $\|\alpha f\| = |\alpha| \, \|f\|$, $\|f\| = 0$ *if and only if* $f = 0$

(ii) $\|f + g\| \leq \|f\| + \|g\|$

for all f, g in $\mathcal{V}$ and α in $\mathbb{C}$. The function $f \to \|f\|$ is called a norm and the value of $\|f\|$ is said to be the norm of f.

Many important examples of inner product and normed spaces recur throughout Fourier analysis.

(1.9) EXAMPLES. (i) $\mathbb{R}^n$ with the usual Euclidean inner product

$$(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} x_i y_i \qquad (\mathbf{x}, \mathbf{y} \in \mathbb{R}^n) \ ;$$

(ii) $\mathbb{C}^n$ with the standard inner product

$$(\mathbf{z}, \mathbf{w}) = \sum_{i=1}^{n} z_i \bar{w}_i \qquad (\mathbf{z}, \mathbf{w} \in \mathbb{C}^n) \ ;$$

(iii) more generally, if I is any index set, let $\ell^2(I)$ be the 'sequences' $\mathbf{z} = \{z_i\}_{i \in I}$ of all complex numbers for which $\sum_{i \in I} |z_i|^2 < \infty$. Then, since

$$\Big| \sum_{i \in I} z_i \bar{w}_i \Big| \leq \Big(\sum_{i \in I} |z_i|^2 \Big)^{1/2} \Big(\sum_{i \in I} |w_i|^2 \Big)^{1/2} ,$$

the function

$$(\mathbf{z}, \mathbf{w}) = \sum_{i \in I} z_i \bar{w}_i \qquad \big(\mathbf{z}, \mathbf{w} \in \ell^2(I)\big)$$

defines an inner product on $\ell^2(I)$. The example of $\mathbb{C}^n$ is the special case when $I = \{1, 2, \ldots, n\}$. We shall usually write ℓ^2_n instead of $\mathbb{C}^n$ and ℓ^2 instead of $\ell^2(\mathbb{Z})$.

(iii) Replacing the exponent 2 by any p, $1 \leq p \leq \infty$, we obtain normed spaces. Denote by $\ell^p(I)$ the vector space of sequences $\mathbf{z} = \{z_i\}_{i \in I}$ of complex numbers z_i for which

$$\sum_{i \in I} |z_i|^p \quad (p \neq \infty) \ , \quad \text{resp.} \quad \sup_{i \in I} |z_i| \qquad (p = \infty)$$

are finite. Then under the respective norms

$$\|\mathbf{z}\|_{\ell^p} = \Big(\sum_{i\in I} |z_i|^p\Big)^{1/p} \qquad (p \neq \infty) \ , \qquad\qquad \|\mathbf{z}\|_{\ell^\infty} = \sup_{i\in I} |z_i| \qquad (p = \infty) \ ,$$

the space $\ell^p(I)$ is a normed space. The inequality

$$\Big(\sum_{i\in I} |z_i + w_i|^p\Big)^{1/p} \leq \Big(\sum_{i\in I} |z_i|^p\Big)^{1/p} + \Big(\sum_{i\in z} |w_i|^p\Big)^{1/p}$$

which establishes the triangle inequality

$$\|\mathbf{z} + \mathbf{w}\|_{\ell^p} \leq \|\mathbf{z}\|_{\ell^p} + \|\mathbf{w}\|_{\ell^p}$$

for $\ell^p(I)$ is usually known as Minkowski's inequality. Unless $p = 2$, $\ell^p(I)$ is not an inner product space, but there is a very useful weak substitute for the Cauchy-Schwarz inequality: if $1/p + 1/q = 1$, then

$$\Big|\sum_{i\in I} z_i \bar{w}_i\Big| \leq \|\mathbf{z}\|_{\ell^p} \|\mathbf{w}\|_{\ell^q} \ . \tag{1.10}$$

This is the so-called Hölder's inequality for sequence spaces.

Through the use of a norm function the standard definitions for convergence of sequences and infinite series of real numbers can be carried over *verbatim* to any normed space replacing the absolute value of a real number by the norm of an element in the normed space.

(1.11) DEFINITION. *Let $\mathcal{V}$ be a normed space. A sequence $\{f_n\}$ in $\mathcal{V}$ is said to be a Cauchy sequence if to each $\varepsilon > 0$ there corresponds N so that*

$$\|f_m - f_n\| < \varepsilon \qquad (m, n > N) \ ,$$

while $\{f_n\}$ is said to be a convergent sequence and have limit $\lim_{n\to\infty} f_n = f$, $f \in \mathcal{V}$, if to each $\varepsilon > 0$ there corresponds N so that

$$\|f_n - f\| < \varepsilon \qquad (n > N) \ .$$

Every convergent sequence is automatically a Cauchy sequence, but there exist normed spaces containing Cauchy sequences which are not convergent sequences. For example, denote by $\mathcal{L}^p(\mathbb{R})$ the space of continuous compactly supported functions $f = f(x)$ on $\mathbb{R}$ with norm

$$\|f\|_p = \left(\int_{-\infty}^{\infty} |f(x)|^p \, dx\right)^{1/p} .$$

It is not hard to construct a sequence of continuous compactly supported functions $\{f_n\}$ such that

(i) *$\{f_n\}$ is a Cauchy sequence in $\mathcal{L}^p(\mathbb{R})$,*

(ii) *f_n converges to the characteristic function $\chi = \chi_{[-\frac{1}{2},\frac{1}{2})}$ of the interval $[-\frac{1}{2},\frac{1}{2})$ in the sense that*

$$\lim_{n\to\infty} \left(\int_{-\infty}^{\infty} |f_n(x) - \chi(x)|^p \, dx\right)^{1/p} = 0 .$$

Since χ is not continuous, $\mathcal{L}^p(\mathbb{R})$ thus contains Cauchy sequences that are not convergent sequences. On the contrary, every Cauchy sequence of real numbers converges to a real number — the *completeness property of* $\mathbb{R}$ — a fact which is fundamental to all of analysis. The corresponding property for a normed space is equally fundamental.

(1.12) Definition. *A normed spaced $\mathfrak{X}$ is said to be a complete normed space or, frequently, a Banach space when every Cauchy sequence in $\mathfrak{X}$ is a convergent sequence. A complete inner product space is usually called a Hilbert space (and denoted here by $\mathcal{H}, \mathcal{H}_0, \mathcal{H}_1, \ldots$).*

The completeness property of $\mathbb{R}$ ensures that every $\ell^p(I)$-space is a Banach space, while $\ell^2(I)$ is a Hilbert space. On the other hand, after replacing Riemann integrals by Lebesgue integrals in our definition of $\mathcal{L}^p$-spaces we obtain the corresponding family of Banach spaces on $\mathbb{R}$. Thus, let $L^p(\mathbb{R})$, $(1 \le p < \infty)$ be the usual space of (equivalence classes) of Lebesgue measurable functions f on $\mathbb{R}$ for which

(1.13)(i) $$\|f\|_{L^p} = \left(\int_{-\infty}^{\infty} |f(x)|^p \, dx\right)^{1/p}$$

is finite. Then $L^p(\mathbb{R})$ is a Banach space containing $\mathcal{L}^p(\mathbb{R})$ as a dense subspace, i.e., to each f in $L^p(\mathbb{R})$ and each $\varepsilon > 0$ there corresponds g in $\mathcal{L}^p(\mathbb{R})$ so that

$$\text{(1.13)(ii)} \qquad \|f-g\|_{L^p} = \left(\int_{-\infty}^{\infty} |f(x)-g(x)|^p\,dx\right)^{1/p} < \varepsilon\ .$$

We shall refer to this property as the $\mathcal{L}^p(\mathbb{R})$-density property of $L^p(\mathbb{R})$. Most crucially of all, the Lebesgue $L^2(\mathbb{R})$-space is a Hilbert space with respect to the inner product

$$\text{(1.14)} \qquad (f,g) = \int_{-\infty}^{\infty} f(x)\,\overline{g(x)}\,dx \qquad \big(f,g\in L^2(\mathbb{R})\big)\ ,$$

and $L^2(\mathbb{R})$ contains $\mathcal{L}^2(\mathbb{R})$ as a dense subspace. The Cauchy-Schwarz inequality for $L^2(\mathbb{R})$ ensures that

$$\text{(1.15)(i)} \qquad \left|\int_{-\infty}^{\infty} f(x)\,\overline{g(x)}\,dx\right| \le \left(\int_{-\infty}^{\infty} |f(x)|^2\,dx\right)^{1/2}\left(\int_{-\infty}^{\infty} |g(x)|^2\,dx\right)^{1/2}.$$

The substitute for Lebesgue $L^p(\mathbb{R})$-spaces is an integral form of Hölder's inequality: *when* $1/p+1/q=1$, *the inequality*

$$\text{(1.15)(ii)} \qquad \left|\int_{-\infty}^{\infty} f(x)\,\overline{g(x)}\,dx\right| \le \left(\int_{-\infty}^{\infty} |f(x)|^p\,dx\right)^{1/p}\left(\int_{-\infty}^{\infty} |g(x)|^q\,dx\right)^{1/q}$$

holds for all f in $L^p(\mathbb{R})$ and g in $L^q(\mathbb{R})$.

Two properties of Hilbert spaces that make essential use of the inner product structure and the completeness property are natural analogues of familiar geometric results in 3-space. Let

$$\mathbf{a} = (a_1,a_2,a_3)\ , \quad \mathbf{b} = (b_1,b_2,b_3)\ , \quad \mathbf{c} = (c_1,c_2,c_3)$$

be vectors in $\mathbb{R}^3$. Then $\{\mathbf{a},\mathbf{b},\mathbf{c}\}$ is a linearly independent family, hence a basis for $\mathbb{R}^3$, if and only if $\mathbf{a}.(\mathbf{b}\times\mathbf{c}) \ne 0$ or, equivalently,

$$\det\begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix} \ne 0\ ;$$

geometrically, $\mathbf{a}.(\mathbf{b}\times\mathbf{c})$ is the volume of the parallelepiped having $\mathbf{a},\mathbf{b}$ and $\mathbf{c}$ as adjacent edges. Cramer's rule says, essentially, that every $\mathbf{v} = (v_1,v_2,v_3)$ in $\mathbb{R}^3$ can be written uniquely as

$$\text{(1.16)} \qquad \mathbf{v} = \alpha\mathbf{a} + \beta\mathbf{b} + \gamma\mathbf{c}$$

with

$$\alpha = \frac{\mathbf{v}.(\mathbf{b}\times\mathbf{c})}{\mathbf{a}.(\mathbf{b}\times\mathbf{c})}\ ,\qquad \beta = \frac{\mathbf{v}.(\mathbf{c}\times\mathbf{a})}{\mathbf{a}.(\mathbf{b}\times\mathbf{c})}\ ,\qquad \gamma = \frac{\mathbf{v}.(\mathbf{a}\times\mathbf{b})}{\mathbf{a}.(\mathbf{b}\times\mathbf{c})}$$

whenever $\mathbf{a}.(\mathbf{b}\times\mathbf{c}) \neq 0$. Now let us consider the special case when $\mathbf{c} = \mathbf{a}\times\mathbf{b}$ so that

$$\mathbf{v} = \alpha\mathbf{a} + \beta\mathbf{b} + \left(\frac{\mathbf{v}.(\mathbf{a}\times\mathbf{b})}{\|\mathbf{a}\times\mathbf{b}\|^2}\right)(\mathbf{a}\times\mathbf{b}) = \mathbf{u} + \mathbf{w}$$

where $\mathbf{u}$ lies in the plane containing $\mathbf{a}, \mathbf{b}$ and $\mathbf{w}$ is perpendicular to this plane. Geometrically, $\|\mathbf{w}\|$ is the shortest distance from the point $\mathbf{v} = (v_1, v_2, v_3)$ in $\mathbb{R}^3$ to the plane containing $\mathbf{a}, \mathbf{b}$. (Where do these properties use the completeness property of $\mathbb{R}$?) To extend such geometric ideas to an arbitrary Hilbert space $\mathcal{H}$, let $\mathcal{V}$ be a closed subspace of $\mathcal{H}$ and f an element of $\mathcal{H}$ not lying in $\mathcal{V}$, and set

$$d = \inf\{\|f - g\|_{\mathcal{H}} : g \in \mathcal{V}\}\ ;$$

the value of d might well be called the *distance* from f to $\mathcal{V}$.

(1.17) THEOREM. *There exist unique elements g, h in $\mathcal{H}$ such that*

(i) $f = g + h$, (ii) $g \in \mathcal{V}$,

(iii) $(h, v) = 0 \quad (v \in \mathcal{V})$, (iv) $\|h\| = d$.

Proof. To construct g set

$$C_n = \left\{v \in \mathcal{V} : \|f - v\|_{\mathcal{H}} \leq d + \frac{1}{n}\right\}\ .$$

In view of the definition of d, each C_n is non-empty and $C_n \supseteq C_{n+1}$. We use the parallelogram law and the completeness property to show that $\bigcap_n C_n$ contains precisely one element — the element g.

For arbitrary u, v in C_n

$$\begin{aligned}\|u - v\|^2 &= \|(u - f) + (f - v)\|^2\\ &= 2(\|u - f\|^2 + \|v - f\|^2) - \|(u - f) - (f - v)\|^2\\ &= 2(\|u - f\|^2 + \|v - f\|^2) - 4\|\tfrac{1}{2}(u + v) - f\|^2\end{aligned}$$

by the parallelogram law. But

$$\|u - f\|^2 + \|v - f\|^2 \leq 2\Big(d + \frac{1}{n}\Big)^2 \qquad (u, v \in C_n)\ ,$$

while

$$\|f - \tfrac{1}{2}(u+v)\| \geq d \ .$$

Consequently,

$$\|u - v\|^2 \leq \frac{8}{n} + \frac{4}{n^2} \qquad (u, v \in C_n) \ ,$$

and so

$$(*) \qquad \lim_{n\to\infty} \text{ diameter } (C_n) = 0 \ .$$

Now choose any g_n in C_n. Since $C_n \supseteq C_{n+1} \supseteq \cdots$, property $(*)$ ensures that $\{g_n\}$ is a Cauchy sequence, which must therefore converge to some g in $\mathcal{V}$. On the other hand, if $\{g'_n\}$ is any other choice of sequence, property $(*)$ ensures also that

$$\lim_{n\to\infty} \|g_n - g'_n\| = 0 \ .$$

Hence $\bigcap_n C_n = \{g\}$.

Set $h = f - g$. Then $\|h\| = d$, since

$$d \leq \|f - g\| \leq \|f - g_n\| + \|g - g_n\| \leq d + \frac{1}{n} + \|g - g_n\| \to d \ ,$$

establishing properties (i), (ii) and (iv). To establish (iii), it is sufficient to show that

$$(**) \qquad (h, v) = 0 \qquad (v \in \mathcal{V} \ , \ \|v\| = 1) \ .$$

But, for any v in $\mathcal{V}$ with $\|v\| = 1$,

$$\begin{aligned} d^2 &\leq \|f - \big(g + (h,v)v\big)\|^2 = \|h - (h,v)v\|^2 \\ &= \|h\|^2 - |(h,v)|^2 \leq d^2 - |(h,v)|^2 \ , \end{aligned}$$

by the same argument as we used in the proof of (1.7)(i). Hence $(h, v) = 0$, establishing $(**)$.

To complete the proof, therefore, we have only to show that g and h are unique. But if $f = g' + h'$ is a second decomposition of f satisfying (ii) and (iv), then $g' \in \bigcap_n C_n$; and so $g = g'$. It follows automatically that $h = h'$, completing the proof. ∎

Just as non-zero vectors $\mathbf{u}, \mathbf{v}$ in 3-space are orthogonal if and only if $\mathbf{u}.\mathbf{v} = 0$, so we say that non-zero elements f, g in an inner product space are *orthogonal* when $(f, g) = 0$. With this terminology we obtain a very useful corollary to the previous result.

(1.18) COROLLARY. *Each closed subspace $\mathcal{V}$ of a Hilbert space $\mathcal{H}$ admits an orthogonal complement*

$$\mathcal{V}^{\perp} = \{u \in \mathcal{H} : (u, v) = 0 \ , \ v \in \mathcal{V}\}$$

in the sense that $\mathcal{H} = \mathcal{V}^{\perp} \oplus \mathcal{V}$, i.e., every f in $\mathcal{H}$ can be written uniquely in the form

$$f = u + v \qquad (u \in \mathcal{V}^{\perp} \ , \ v \in \mathcal{V}) \ .$$

We turn now to the theory associated with particular families of elements in a Hilbert space, the first of which is the natural abstraction of the vectors

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

in $\mathbb{R}^3$. Geometrically, these are mutually orthogonal vectors in $\mathbb{R}^3$ having unit length since they have the properties

$$||\mathbf{i}|| = ||\mathbf{j}|| = ||\mathbf{k}|| = 1 \ , \qquad \mathbf{i}.\mathbf{j} = \mathbf{j}.\mathbf{k} = \mathbf{k}.\mathbf{i} = 0$$

with respect to the inner product (1.2) on 3-space.

(1.19) DEFINITION. *A subset $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots\}$ of an inner product space $\mathcal{V}$ is said to be an orthonormal family when*

$$||\mathbf{v}_j|| = 1 \ , \quad (\mathbf{v}_j, \mathbf{v}_k) = 0 \qquad (j \neq k) \ ,$$

i.e., S consists of mutually orthogonal elements having length 1.

For example, the set $\{\mathbf{e}_i : i \in I\}$ where

$$\mathbf{e}_i = (0, 0, \ldots, 0, 1, 0, \ldots)$$

is the sequence having 0 entries except at the i^{th} place, is orthonormal in $\ell^2(I)$. As frequently happens, however, a subset S of an inner product space $\mathcal{V}$ can consist of linearly independent elements which are not mutually orthogonal. In such a case there are several ways of constructing an orthonormal family which retains a particular geometric or algebraic property that S has. One method is very familiar.

Let $S = \{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$ be the linearly independent set of vectors in 3-space in (1.16). Then $\mathbf{a}$ and $\mathbf{b}$ are not parallel, so there is a unique plane in which $\mathbf{a}, \mathbf{b}$ lie, and the vector $\mathbf{c}$ cannot lie in the plane (why?). Now normalize $\mathbf{a}$ by setting $\boldsymbol{\alpha} = \mathbf{a}/\|\mathbf{a}\|$, and define $\boldsymbol{\beta}$ by

$$\boldsymbol{\beta} = \big(\mathbf{b} - (\mathbf{b}.\,\boldsymbol{\alpha})\,\boldsymbol{\alpha}\big)\Big/\|\mathbf{b} - (\mathbf{b}.\,\boldsymbol{\alpha})\,\boldsymbol{\alpha}\,\| \ .$$

Then $\boldsymbol{\beta}$ is a unit vector in the plane containing $\mathbf{a}, \mathbf{b}$, and $\boldsymbol{\beta}$ is orthogonal to $\boldsymbol{\alpha}$ since

$$\boldsymbol{\beta}\,.\,\boldsymbol{\alpha} = \big(\mathbf{b}.\,\boldsymbol{\alpha} - (\mathbf{b}.\,\boldsymbol{\alpha})\|\,\boldsymbol{\alpha}\,\|^2\big)\Big/\|\mathbf{b} - (\mathbf{b}.\,\boldsymbol{\alpha})\,\boldsymbol{\alpha}\,\| = 0 \ .$$

Finally, to construct a unit vector $\boldsymbol{\gamma}$ perpendicular to the plane containing $\mathbf{a}, \mathbf{b}$, and hence to the plane containing $\boldsymbol{\alpha}, \boldsymbol{\beta}$, we can take for $\boldsymbol{\gamma}$ the normalized *projection* of $\mathbf{c}$ onto the vector normal to this plane. As $\mathbf{c}$ cannot lie in this plane, its projection onto the normal must be non-zero. In detail:

$$\boldsymbol{\gamma} = \big(\mathbf{c} - (\mathbf{c}.\,\boldsymbol{\alpha})\,\boldsymbol{\alpha} - (\mathbf{c}.\,\boldsymbol{\beta})\,\boldsymbol{\beta}\big)\Big/\|\mathbf{c} - (\mathbf{c}.\,\boldsymbol{\alpha})\,\boldsymbol{\alpha} - (\mathbf{c}.\beta)\,\boldsymbol{\beta}\,\| \ .$$

An explicit computation shows that $\boldsymbol{\gamma}\,.\,\boldsymbol{\alpha} = \boldsymbol{\gamma}\,.\,\boldsymbol{\beta} = 0$. Hence $\{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}\}$ is an orthonormal family in 3-space. This construction carries over almost immediately to any inner product space; it is known as the *Gram-Schmidt orthogonalization process.*

(1.20) THEOREM. (Gram-Schmidt process). *If $\{\beta_1, \beta_2, \ldots\}$ is a possibly infinite set of linearly independent elements in an inner product space $\mathcal{V}$, then*

$$\phi_1 = \beta_1/\|\beta_1\| \ , \qquad \phi_k = \Big(\beta_k - \sum_{j=1}^{k-1}(\beta_k, \phi_j)\phi_j\Big)\Big/\Big\|\beta_k - \sum_{j=1}^{k-1}(\beta_k, \phi_j)\phi_j\Big\| \qquad (k > 1)$$

defines a family $S = \{\phi_1, \phi_2, \ldots\}$ which is orthonormal in $\mathcal{V}$.

Let us now return to the case of a general orthonormal family $S = \{\phi_1, \phi_2, \ldots\}$ in an inner product space $\mathcal{V}$ having inner product $(.,.)$. To each f in $\mathcal{V}$ there corresponds an *orthonormal series*

$$S[f] = \sum_k (f, \phi_k)\phi_k \qquad (\phi \in \mathcal{V}) \ ,$$

and the n^{th} *partial sum* of this series is given by

$$S_n[f] = \sum_{k=1}^{n} (f, \phi_k)\phi_k \qquad (\phi \in \mathcal{V}) \ .$$

The coefficients (f, ϕ_k) of such a partial sum have a very special significance in approximation theory indicating why they and the associated orthonormal series will play such an important role in the future. Let $a_1, \ldots, a_n$ be arbitrary scalars. Then, because of the orthonormality of the ϕ_k,

$$\begin{aligned} \Big\| f - \sum_{k=1}^{n} a_k\phi_k \Big\|^2 &= \Big(f - \sum_{j=1}^{n} a_j\phi_j \ , \ f - \sum_{k=1}^{n} a_k\phi_k \Big) \\ &= \|f\|^2 - \sum_{j=1}^{n} a_j(\phi_j, f) - \sum_{k=1}^{n} \bar{a}_k(f, \phi_k) + \sum_{k=1}^{n} |a_k|^2 \qquad (1.21) \\ &= \|f\|^2 - \sum_{k=1}^{n} |(f, \phi_k)|^2 + \sum_{k=1}^{n} |a_k - (f, \phi_k)|^2 \ , \end{aligned}$$

using the symmetry and linearity properties of the inner product on $\mathcal{V}$. Thus

$$\Big\| f - \sum_{k=1}^{n} a_k\phi_k \Big\| \qquad (1.22)$$

is minimized by $\sum_{k=1}^{n} |(f, \phi_k)|^2$ when $a_k = (f, \phi_k)$; in fact, any other choice of the a_k would give a strictly larger value to (1.22). Hence the n^{th} partial sum $\sum_{k=1}^{n}(f, \phi_k)\phi_k$ provides the *best approximation* to f by a linear combination of $\phi_1, \ldots, \phi_n$, distance being measured in terms of the length $\|.\|$. In technical terms, the n^{th} partial sum $S_n[f]$ is said to provide the *least squares approximation* to f by linear combinations of $\phi_1, \ldots, \phi_n$.

(1.23) THEOREM. *Let $\{\phi_1, \phi_2, \ldots\}$ be an orthonormal family in an inner product space $\mathcal{V}$. Then*

(i) (Bessel's equality)

$$\Big\| f - \sum_{k=1}^{n} (f, \phi_k)\phi_k \Big\|^2 = \|f\|^2 - \sum_{k=1}^{n} |(f, \phi_k)|^2 \ ,$$

(ii) (Bessel's inequality)

$$\left(\sum_{k=1}^{n} |(f, \phi_k)|^2\right)^{1/2} \leq \|f\|$$

hold for each n

Bessel's equality follows from (1.21), and Bessel's inequality then follows immediately from his equality since $\|.\|$ is always non-negative. This leads us to another crucial point in our development of Hilbert spaces. Suppose $\mathcal{V}$ has finite dimension, say dimension n. Then $\mathcal{V}$ contains a linearly independent set $\{\phi_1, \ldots, \phi_n\}$ which because of (1.20) can be taken to be orthonormal, and by general vector space theory each f has a *unique* expansion

$$f = a_1 f_1 + a_2 f_2 + \cdots + a_n f_n \ .$$

In this case, the orthonormality of the ϕ_k ensures that

$$(f, \phi_k) = \sum_{j=1}^{n} a_j (f, \phi_k) = a_k \qquad (1 \leq k \leq n) \ .$$

Hence

(1.24)(i) $$f = (f, \phi_1)\phi_1 + (f, \phi_2)\phi_2 + \cdots + (f, \phi_n)\phi_n \ ,$$

and so by Bessel's equality,

(1.24)(ii) $$\left(\sum_{k=1}^{n} |(f, \phi_k)|^2\right)^{1/2} = \|f\| \qquad (f \in \mathcal{V}) \ ;$$

in particular, $\|f\|$ coincides with the length in ℓ_n^2 of the n-tuple $((f, \phi_1), \ldots, (f, \phi_n))$ of coefficients of f. Conversely, to each $\mathbf{a} = (a_1, \ldots, a_n)$ in ℓ_n^2 there corresponds an element

(1.25)(i) $$f = a_1\phi_1 + a_2\phi_2 + \cdots + a_n\phi_n$$

such that

(1.25)(ii) $$a_k = (f, \phi_k) \ , \qquad \|f\| = \|\mathbf{a}\| = \left(\sum_{k=1}^{n} |a_k|^2\right)^{1/2} \ .$$

Thus, through (1.24) and (1.25), $f \leftrightarrow \mathbf{a}$ defines a length-preserving isomorphism between $\mathcal{V}$ and ℓ_n^2; consequently, when $\mathcal{V}$ has dimension n, the orthonormal family $\{\phi_1, \ldots, \phi_n\}$ behaves in every respect like the standard family $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ in 3-space. Now suppose $\mathcal{V}$ has infinite dimension and let $\{\phi_1, \phi_2, \ldots\}$ be an infinite orthonormal family in $\mathcal{V}$. Bessel's inequality still holds for every n; consequently, after letting $n \to \infty$ we obtain the infinite version of (1.23)(ii).

(1.26) COROLLARY. (Bessel's inequality) *If* $\{\phi_1, \phi_2, \ldots\}$ *is an infinite orthonormal family, then*

$$\left(\sum_{k=1}^{\infty} |(f, \phi_k)|^2\right)^{1/2} \leq \|f\|$$

is valid for all f *in* $\mathcal{V}$*; in particular, the sequence* $\{(f, \phi_k)\}_{k=1}^{\infty}$ *of coefficients is always in* ℓ^2.

But, conversely, given any sequence $\mathbf{a} = \{a_k\}$ in ℓ^2, what can be said about the *infinite* series $\sum_{k=1}^{\infty} a_k \phi_k$? Can sense be made of the convergence of it? If so, is the sum an element of $\mathcal{V}$? As with theorem (1.17), the completeness of the inner product space is essential in answering these questions. Suppose first that $\{a_k\}_{k=1}^{\infty}$ is an arbitrary element of ℓ^2, and set $g_n = \sum_{k=1}^{n} a_k \phi_k$. Then to each $\varepsilon > 0$ there corresponds an integer N such that

$$|a_{m+1}|^2 + \cdots + |a_n|^2 = \sum_{k=m+1}^{n} |a_k|^2 < \varepsilon$$

for all $m, n > N$; consequently, by Bessel's inequality,

$$\|g_n - g_m\|^2 = \left\| \sum_{k=m+1}^{n} a_k \phi_k \right\|^2 = \sum_{k=m+1}^{n} |a_k|^2 < \varepsilon$$

for all $m, n > N$. Thus $\{g_n\}$ is a Cauchy sequence in $\mathcal{V}$. If $\mathcal{V}$ is assumed to be complete, i.e., a Hilbert space rather than just an inner product space, this Cauchy sequence $\{g_n\}$ will converge to some g in $\mathcal{V}$, i.e.,

$$g = \lim_{n \to \infty} \sum_{k=1}^{n} a_k \phi_k \ ;$$

furthermore,

$$(g, \phi_j) = \lim_{n \to \infty} \left(\sum_{k=1}^{n} a_k \phi_k, \phi_j \right) = a_j \ .$$

Consequently, if $\{\phi_k\}$ is an orthonormal family in a Hilbert space, then $\sum_k a_k \phi_k$ converges to some g such that

$$g = \sum_k a_k \phi_k = \sum_k (g, \phi_k) \phi_k$$

for each element $\{a_k\}$ in ℓ^2. The crucial question now is whether every element of the Hilbert space can be written as $\sum_k a_k \phi_k$ for some $\{a_k\}$ in ℓ^2.

(1.27) Definition. *An infinite family* $\{\phi_1, \phi_2, \ldots\}$ *in a (separable) Hilbert space* $\mathcal{H}$ *is said to be a complete orthonormal family if for each* f *in* $\mathcal{H}$

$$\lim_{n\to\infty} \left\| f - \sum_{k=1}^{n} (f, \phi_k)\phi_k \right\| = 0 \ .$$

In such a case the orthonormal series $\sum_{k=1}^{\infty} (f, \phi_k)\phi_k$ *is said to converge in mean to* f.

The most basic consequence of the existence of such complete orthonormal families is contained in the next result.

(1.28) Theorem. *For an infinite family* $\{\phi_1, \phi_2, \ldots\}$ *in a (separable) Hilbert space* $\mathcal{H}$ *the following assertions are equivalent:*

(i) $\{\phi_1, \phi_2, \ldots\}$ *is a complete orthonormal family in* $\mathcal{H}$,

(ii) *(Plancherel)* $\quad \|f\|^2 = \sum_{k=1}^{\infty} |(f, \phi_k)|^2 \qquad (f \in \mathcal{H})$,

(iii) *(Parseval)* $\quad (f, g) = \sum_{k=1}^{\infty} (f, \phi_k)\overline{(g, \phi_k)} \qquad (f, g \in \mathcal{H})$,

(iv) *if* $(f, \phi_k) = 0$ *for all* $k = 1, 2, \ldots$, *then* $f = \mathbf{0}$.

Proof. (i) $\Rightarrow$ (ii) In view of Bessel's inequality, the complete orthonormal family property ensures that

$$\lim_{n\to\infty} \left(\|f\|^2 - \sum_{k=1}^{n} |(f, \phi_k)|^2 \right) = \lim_{n\to\infty} \left\| f - \sum_{k=1}^{n} (f, \phi_k)\phi_k \right\|^2 = 0 \ .$$

Thus the equality

$$\|f\|^2 = \sum_{k=1}^{\infty} |(f, \phi_k)|^2$$

holds for each $f \in \mathcal{H}$.

(ii) $\Rightarrow$ (iii) We apply the polarization identity (1.7)(iv). For then by Plancherel's theorem,

$$\begin{aligned}
(f, g) &= \tfrac{1}{4}\{\|f+g\|^2 - \|f-g\|^2 + i\|f+ig\|^2 - i\|f-ig\|^2\} \\
&= \tfrac{1}{4}\sum_k \{|(f+g, \phi_k)|^2 - |(f-g, \phi_k)|^2 + i|(f+ig, \phi_k)|^2 - i(f-ig, \phi_k)|^2\} \\
&= \tfrac{1}{4}\sum_k \{|(f, \phi_k) + (g, \phi_k)|^2 - |(f, \phi_k) - (g, \phi_k)|^2 + i|(f, \phi_k) + i(g, \phi_k)|^2 - i|(f, \phi_k) - i(g, \phi_k)|^2\} \\
&= \sum_k (f, \phi_k)\overline{(g, \phi_k)} \ ,
\end{aligned}$$

applying the polarization identity to the Hilbert space $\mathbb{C}$ also.

(iii) $\Rightarrow$ (iv) If $(f, \phi_k) = 0$ for all $k = 1, 2, \ldots$, then

$$(f, f) = \sum_k |(f, \phi_k)|^2 = 0 \ ,$$

i.e., $\|f\| = 0$. So $F = 0$.

(iv) $\Rightarrow$ (i) Let g be an arbitrary element of $\mathcal{H}$. Then by (1.26) the sequence $\{g, \phi_k)\}_k$ of coefficients is in ℓ^2, and so by the discussion prior to definition (1.27), the series $\sum_k (g, \phi_k)\phi_k$ converges to some h in $\mathcal{H}$ such that $(h, \phi_k) = (g, \phi_k)$. Now let's apply condition (iv) to $f = g - h$; thus $g = h$, i.e., $g = \sum_{k=1}^{\infty}(g, \phi_k)\phi_k$. But then

$$0 = \lim_{n \to \infty} \|h - g_n\| = \lim_{h \to \infty} \|g - g_n\|$$

showing that $\{\phi_1, \phi_2, \ldots\}$ is a *complete* orthonormal family in $\mathcal{H}$. ∎

Consequently, if $\{\phi_i\}_{i \in I}$ is a complete orthonormal family in a Hilbert space $\mathcal{H}$, then every f has an orthonormal series decomposition $f = \sum_i (f, \phi_i)\phi_i$ such that $\|f\| = \sum_i |(f, \phi_i)|^2$, and this is the only possible decomposition of f as a sum of the ϕ_i. But in developing the theory of wavelets one often starts with a family that is not orthonormal. Now of course, the Gram-Schmidt process could be applied to produce an orthonormal family, but this destroys the time-scale properties that are basic to wavelet decompositions. So the notion of orthonormality has to be weakened while still preserving the unique decomposition property. In the next section we shall weaken the notion still further, eliminating even the unique decomposition requirement.

(1.29) Definition. *A family $\{\phi_i\}_{i \in I}$ in a Hilbert space $\mathcal{H}$ is said to be stable when there exist positive constants A, B so that the inequalities*

$$A\Big(\sum_i |\alpha_i|^2\Big)^{1/2} \le \Big\|\sum_i \alpha_i \phi_i\Big\|_{\mathcal{H}} \le B\Big(\sum_i |\alpha_i|^2\Big)^{1/2}$$

hold for all $\boldsymbol{\alpha} = \{\alpha_i\}$ in $\ell^2(I)$. The constants A, B will be called ***stability constants****.*

Since $\|\lambda x\|_{\mathcal{H}} = |\lambda|\,\|x\|_{\mathcal{H}}$, $\lambda \in \mathbb{C}$, the stable property is equivalent to the existence of constants A, B so that

$$(2.19)' \qquad 0 < A \le \Big\|\sum_i \alpha_i \phi_i\Big\|_{\mathcal{H}} \le B \qquad \big(\|\boldsymbol{\alpha}\| = 1\big)$$

for all *norm one* $\boldsymbol{\alpha} = \{\alpha_i\}$ in $\ell^2(I)$, i.e., all $\boldsymbol{\alpha}$ in the unit sphere of $\ell^2(I)$.

Important examples of stable families will arise in the theory of wavelets, but there are examples of a more general nature.

(1.30) EXAMPLES. (i) Every orthonormal family is stable with stability constants $A = B = 1$.

(ii) Every finite linearly independent set is stable. Indeed, for any finite set $\phi_1, \ldots, \phi_n$ in a Hilbert space $\mathcal{H}$, define $f : \mathbb{C}^n \to [0, \infty)$ by

$$f(\boldsymbol{\alpha}) = \|\alpha_1\phi_1 + \cdots + \alpha_n\phi_n\| \qquad (\boldsymbol{\alpha} \in \mathbb{C}^n) \ .$$

Then f is continuous and so f is bounded above and below on the unit sphere in $\mathbb{C}^n$, i.e.

$$0 \le A \le \Big\| \sum_{i=1}^{n} \alpha_i\phi_i \Big\| \le B \qquad (\|\boldsymbol{\alpha}\| = 1)$$

where

$$A = \min_{\|\boldsymbol{\alpha}\|=1} f(\boldsymbol{\alpha}) \quad , \quad B = \max_{\|\boldsymbol{\alpha}\|=1} f(\boldsymbol{\alpha}) \ .$$

But, if $\{\phi_1, \ldots, \phi_n\}$ is linearly independent, $f(\boldsymbol{\alpha})$ vanishes only at $\boldsymbol{\alpha} = 0$, ensuring that $A > 0$. Hence every finite set of linearly independent elements in a Hilbert space is a stable family.

(iii) Not every infinite set is stable but there are infinite non-orthonormal stable families. For if $\{\phi_1, \phi_2, \ldots\}$ is orthonormal, define $\{\psi_1, \psi_2, \ldots\}$ by

$$\begin{aligned} \psi_1 &= \frac{1}{\sqrt{2}}(\phi_1 + \phi_2) \ , \qquad & \psi_2 &= \frac{1}{\sqrt{2}}(\phi_2 + \phi_3) \ , \\ \psi_3 &= \frac{1}{\sqrt{2}}(\phi_4 + \phi_5) \ , \qquad & \psi_4 &= \frac{1}{\sqrt{2}}(\phi_5 + \phi_6) \ , \\ &\vdots & & . \end{aligned}$$

Then

$$\|\psi_j\| = 1 \ , \qquad (\psi_{2j-1}, \psi_{2j}) = \tfrac{1}{2} \quad (\text{all } j)$$

while $(\psi_j, \psi_k) = 0$ for all other choices of j, k. Thus

$$\Big\| \sum_{i=1}^{\infty} \alpha_i\psi_i \Big\|^2 = \sum_{i=1}^{\infty} |\alpha_i|^2 + \tfrac{1}{2} \sum_{i=1}^{\infty} (\alpha_{2i-1}\bar{\alpha}_{2i} + \bar{\alpha}_{2i-1}\alpha_{2i}) \ ,$$

and so

$$\sum_i \big(|\alpha_i|^2 - |\alpha_{2i-1}\alpha_{2i}|\big) \le \Big\| \sum_i \alpha_i \psi_i \Big\|^2 \le \sum_i \big(|\alpha_i|^2 + |\alpha_{2i-1}\alpha_{2i}|\big) \ .$$

But by the arithmetic-geometric mean inequality

$$|\alpha_{2i-1}\alpha_{2i}| \le \tfrac{1}{2}(|\alpha_{2i-1}|^2 + |\alpha_{2i}|^2) \ .$$

Consequently,

$$\sqrt{\frac{1}{2}}\Big(\sum_i |\alpha_i|^2\Big)^{1/2} \le \Big\| \sum_i \alpha_i \psi_i \Big\| \le \sqrt{\frac{3}{2}}\Big(\sum_i |\alpha_i|^2\Big)^{1/2} \ ,$$

showing that

$$\Big\{ \frac{1}{\sqrt{2}}(\phi_1 + \phi_2), \frac{1}{\sqrt{2}}(\phi_2 + \phi_3), \frac{1}{\sqrt{2}}(\phi_4 + \phi_5), \dots \Big\}$$

is a stable family.

Several elementary Hilbert space properties of stable families will be needed later. So let $\{\phi_i\}_{i \in I}$ be a stable family in $\mathcal{H}$ and set

$$\mathcal{H}_0 = \Big\{ \sum_i \alpha_i \phi_i : \{\alpha_i\} \in \ell^2(I) \Big\} \ . \tag{1.31}$$

For each n choose any $\boldsymbol{\alpha}^{(n)} = \{\alpha_i^{(n)}\}$ in $\ell^2(I)$ and let

$$f_n = \sum_i \alpha_i^{(n)} \phi_i$$

be the corresponding element in $\mathcal{H}_0$. Then by stability,

$$A \| \, \boldsymbol{\alpha}^{(m)} - \boldsymbol{\alpha}^{(n)} \, \|_{\ell^2} \le \|f_m - f_n\| \le B \| \, \boldsymbol{\alpha}^{(m)} - \boldsymbol{\alpha}^{(n)} \, \|_{\ell^2}$$

in which case f_n will be a Cauchy sequence in $\mathcal{H}$ if and only if $\{\boldsymbol{\alpha}^{(n)}\}$ is a Cauchy sequence in $\ell^2(I)$. Since $\ell^2(I)$ is a Hilbert space, it thus follows that $\mathcal{H}_0$ is a closed subspace of $\mathcal{H}$ — intuitively, $\mathcal{H}_0$ is a 'copy' of $\ell^2(I)$ in $\mathcal{H}$.

(1.32) THEOREM. *Let $\{\phi_i\}_{i\in I}$ be a stable family in $\mathcal{H}$ and define the closed subspace $\mathcal{H}_0$ by* (1.31). *Then the following assertions are equivalent:*

(i) $\mathcal{H}_0 = \mathcal{H}$ (ii) $(f, \phi_i) = 0$, $i \in I$, *if and only if* $f = 0$,

(iii) *finite linear combinations of the ϕ_i are dense in $\mathcal{H}$.*

Proof. (i) ⇒ (iii) Property (iii) asserts that to each f in $\mathcal{H}$ and $\varepsilon > 0$ there corresponds a finite linear combination $\alpha_{i_1}\phi_{i_1} + \cdots + \alpha_{i_n}\phi_{i_n}$ of the ϕ_i so that

$$(*) \qquad \|f - (\alpha_{i_1}\phi_{i_1} + \cdots + \alpha_{i_n}\phi_{i_n})\| < \varepsilon\ .$$

But, if $\mathcal{H} = \mathcal{H}_0$, then every f in $\mathcal{H}$ can be written as $f = \sum_i \alpha_i \phi_i$ for some $\{\alpha_i\}$ in $\ell^2(I)$. To obtain $(*)$ choose $J = \{i_1, \ldots, i_n\}$ so that

$$\left(\sum_{i \notin J} |\alpha_i|^2\right)^{1/2} < \varepsilon / B\ .$$

Then

$$\|f - (\alpha_{i_1}\phi_{i_1} + \cdots + \alpha_{i_n}\phi_{i_n})\| = \left\| \sum_{i\notin J} \alpha_i \phi_i \right\| < \varepsilon\ ,$$

establishing $(*)$.

(iii) ⇒ (ii) Choose any f so that $(f, \phi_i) = 0$ for all i in I. But, by hypothesis, $(*)$ holds and so

$$\varepsilon^2 > \left\| f - \sum_i \alpha_{i_j}\phi_{i_j} \right\|^2 = \|f\|^2 + \left\| \sum_j \alpha_{i_j}\phi_{i_j} \right\|^2$$

since $(f, \phi_{i_j}) = 0$. Thus $\|f\| < \varepsilon$ for each $\varepsilon > 0$. Hence $f = 0$.

(ii) ⇒ (i) Whether or not $\mathcal{H}_0 = \mathcal{H}$, $\mathcal{H}_0$ is a closed subspace of $\mathcal{H}$. So if $\mathcal{H}_0 \neq \mathcal{H}$, there exists f in $\mathcal{H}$ with $f \notin \mathcal{H}_0$, and we can apply theorem (1.18). Thus

$$f = g + h\ , \qquad g \in \mathcal{H}_0\ , \quad h \in \mathcal{H}_0^\perp\ ;$$

in particular, $(h, \phi_i) = 0$ for all $i \in I$. But then $h = 0$, which in turn forces f to be an element of $\mathcal{H}_0$. Hence $\mathcal{H} = \mathcal{H}_0$. ■

When $\mathcal{H} = \mathcal{H}_0$ we shall say that $\{\phi_i\}_{i\in I}$ is a *complete stable family*, or a *complete orthonormal family* if in addition (1.20) holds. Since any stable family is automatically linearly independent, a complete stable family is a basis for $\mathcal{H}$ having the important property that

$$(1.33) \qquad f = \sum_i \alpha_i\phi_i \quad , \quad A\Big(\sum_i |\alpha_i|^2\Big)^{1/2} \le \|f\| \le B\Big(\sum_i |\alpha_i|^2\Big)^{1/2} \; .$$

Such a representation for f is particularly easy to describe in the complete orthonormal case; for then

$$(1.34) \qquad f = \sum_i (f,\phi_i)\phi_i \; , \qquad \|f\| = \Big(\sum_i |f,\phi_i)|^2\Big)^{1/2} \; .$$

The simplicity of this algorithm for computing the coefficients in the representation of f is the first indication of why the search for orthonormal families is so important.

Even if the stable family $\{\phi_i\}_{i\in I}$ is neither complete nor orthonormal, all is not lost however.

(1.35) DEFINITION. *A family $\{\tilde\phi_i\}_{i\in I}$ is said to be a bi-orthogonal family associated with a stable family $\{\phi_i\}_{i\in I}$ when*

(i) $\tilde\phi_i \in \mathcal{H}_0$, *(ii)* $(\phi_i, \tilde\phi_i) = 1$, *(iii)* $(\phi_i, \tilde\phi_j) = 0 \quad (i \ne j)$

for all i, j in I.

For example, in $\mathbb{R}^3$

$$(1.36) \qquad \left\{\frac{\mathbf{b}\times\mathbf{c}}{\mathbf{a}.(\mathbf{b}\times\mathbf{c})} \; , \; \frac{\mathbf{c}\times\mathbf{a}}{\mathbf{a}.(\mathbf{b}\times\mathbf{c})} \; , \; \frac{\mathbf{a}\times\mathbf{b}}{\mathbf{a}.(\mathbf{b}\times\mathbf{c})}\right\}$$

is a bi-orthogonal family associated with any set $\{\mathbf{a},\mathbf{b},\mathbf{c}\}$ of linearly independent vectors $\mathbf{a},\mathbf{b}$, and $\mathbf{c}$. But bi-orthogonal families exist in complete generality: choose a matrix $[\gamma_{jk}]$ so that

$$(1.37)(\mathrm{i}) \qquad \sum_k \gamma_{jk}(\phi_k,\phi_i) = \begin{cases} 0 & (i \ne k) \; , \\ 1 & (i = k) \; , \end{cases}$$

and define $\tilde\phi_i$ by

$$(1.37)(\mathrm{ii}) \qquad \tilde\phi_j = \sum_k \gamma_{jk}\phi_k \qquad (j \in I) \; .$$

Then

$$(\phi_i, \tilde{\phi}_j) = \sum_k \bar{\gamma}_{jk}(\phi_i, \phi_k) = \overline{\sum_k \gamma_{jk}(\phi_k, \phi_i)} \ ,$$

from which (1.25)(ii) and (iii) follow immediately using the defining properties (1.37) of $\tilde{\phi}_j$. The analogue of (1.24) also follows easily.

(1.38) THEOREM. *Let $\{\tilde{\phi}_i\}_{i\in I}$ be a bi-orthogonal family associated with a stable family $\{\phi_i\}_{i\in I}$ in a Hilbert space $\mathcal{H}$ and let $\mathcal{H}_0$ be the closed subspace of $\mathcal{H}$ defined by* (1.21). *Then each f in $\mathcal{H}_0$ admits a series representation*

$$f = \sum_{i\in I} (f, \tilde{\phi}_i)\phi_i$$

in which the coefficients satisfy inequalities

$$(1/B)\Big(\sum_{i\in I} |(f, \tilde{\phi}_i)|^2\Big)^{1/2} \le \|f\| \le (1/A)\Big(\sum_{i\in I} |(f, \tilde{\phi}_i)|^2\Big)^{1/2} .$$

If $\{\phi_i\}_{i\in I}$ is a complete stable family therefore, every f in $\mathcal{H}$ will admit such a series representation. Notice also that the decomposition (1.16) of any $\mathbf{v}$ in 3-space is precisely the one given by $\{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$ and its bi-orthogonal family (1.26).

Proof. By definition, each f in $\mathcal{H}_0$ has a series representation $f = \sum_i \alpha_i \phi_i$ in which the coefficients satisfy inequalities

$$A\Big(\sum_i |\alpha_i|^2\Big)^{1/2} \le \|f\| \le B\Big(\sum_i |\alpha_i|^2\Big)^{1/2} .$$

But then, the bi-orthogonality conditions (1.25)(ii),(iii) ensure that

$$(f, \tilde{\phi}_i) = \Big(\sum_i \alpha_i\phi_i, \tilde{\phi}_j\Big) = \sum_i \alpha_i(\phi_i, \tilde{\phi}_i) = \alpha_j \ .$$

Hence $f = \sum_i (f, \tilde{\phi}_j)\phi_j$ and

$$A\Big(\sum_i |(f, \phi_i)|^2\Big)^{1/2} \le \|f\| \le B\Big(\sum_i |(f, \phi_i)|^2\Big)^{1/2} ,$$

completing the proof. ∎

The following result can be established also.

(1.39) COROLLARY. *A bi-orthogonal family $\{\tilde{\phi}_i\}_{i\in I}$ associated with a stable family $\{\phi_i\}_{i\in I}$ satisfies inequalities*

$$(1/B)\Big(\sum_i |\alpha_i|^2\Big)^{1/2} \leq \Big\|\sum_i \alpha_i\tilde{\phi}_i\Big\| \leq (1/A)\Big(\sum_i |\alpha_i|^2\Big)^{1/2} ;$$

in particular, $\{\tilde{\phi}_i\}_{i\in I}$ is itself a stable family.

In practice, calculation of the γ_{jk} is usually far from easy. Fortunately, when stable families arise in the theory of wavelets, the calculations are reasonably easy and involve only routine but very natural ideas from Fourier analysis as we shall see in section (v) later.

To illustrate the ideas involved in this section we look at some simple though important examples.

The first example is the Discrete Fourier transform (DFT). Let

$$\omega = \exp\left(\frac{2\pi i}{n}\right) = \cos\frac{2\pi}{n} + i\sin\frac{2\pi}{n} .$$

Now on the one hand the family $\{\varepsilon_0, \ldots, \varepsilon_{n-1}\}$, where

$$\varepsilon_i = (0, \ldots, 0, 1, 0, \ldots, 0)$$

is the sequence having zero entries except for the $(i+1)^{st}$ entry, is an orthonormal basis for ℓ_n^2, while on the other hand, $\{\boldsymbol{\xi}_0, \ldots, \boldsymbol{\xi}_{n-1}\}$ also is an orthonormal basis for ℓ_n^2 when

$$\boldsymbol{\xi}_k = \frac{1}{\sqrt{n}}(1, \omega^k, \omega^{2k}, \ldots, \omega^{(n-1)k})$$

since

$$1 + \omega^r + \omega^{2r} + \cdots + \omega^{(n-1)r} = \frac{1-\omega^n}{1-\omega^r} = 0$$

for any integer r, $r \neq 0$. Consequently each z in ℓ_n^2 admits two representations: the first in terms of the ε_i,

$$\mathbf{z} = (z_0, z_1, \ldots, z_{n-1}) = \sum_{i=0}^{n-1} z_i\,\varepsilon_i ,$$

and the second in terms of the $\boldsymbol{\xi}_k$

$$\mathbf{z} = (\hat{z}_0, \hat{z}_1, \ldots, \hat{z}_{n-1}) = \sum_{k=0}^{n-1} \hat{z}_k \, \boldsymbol{\xi}_k \ .$$

Thus

$$\sum_{i=0}^{n-1} z_i \, \boldsymbol{\varepsilon}_i = \sum_{k=0}^{n-1} \hat{z}_k \, \boldsymbol{\xi}_k \ .$$

Clearly

$$\hat{z}_k = \frac{1}{\sqrt{n}} \sum_{j=0}^{n-1} z_j \bar{\omega}^{jk} \tag{1.40}$$

while

$$z_j = \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} \hat{z}_k \omega^{jk} \ . \tag{1.41}$$

The mapping $(z_0, \ldots, z_{n-1}) \to (\hat{z}_0, \ldots, \hat{z}_{n-1})$ is known as the Discrete Fourier transform, and $(\hat{z}_0, \ldots, \hat{z}_{n-1})$ is called the Inverse Discrete Fourier transform.

An interesting and extremely simple orthonormal system in $L^2(\mathbb{R})$ consisting solely of step functions was introduced by Haar in his inaugural dissertation (Göttingen, 1909). Let $h = h(x)$ be the step function defined on $[0, 1)$ by

$$h(x) = \begin{cases} 1 & (0 \leq x < 1/2) \ , \\ -1 & (1/2 \leq x < 1) \ , \end{cases} \tag{1.42}$$

and zero elsewhere. In the original construction this function was used to generate an orthonormal family in $L^2[0, 1]$, but it is in the construction of an orthonormal family in $L^2(-\infty, \infty)$ that Haar functions become the prototype for wavelets in general.

A real number x is said to be a *dyadic number* when $x = k/2^m$ with k, m arbitrary integers (positive, negative or 0), and an interval $[k/2^m, \ (k+1)/2^m)$ is called a *basic dyadic interval*. The fundamental geometric property of such intervals is that two basic dyadic intervals are either disjoint or have the property that one contains the other. This can be proved algebraically, but is most easily established geometrically using the dyadic decomposition of the upper half-plane in *figure* 1: draw the line $y = 2^{-m}$, $-\infty < m < \infty$, parallel to the x-axis and then mark off successively the points $k2^{-m}$, $-\infty < k < \infty$ on this line.

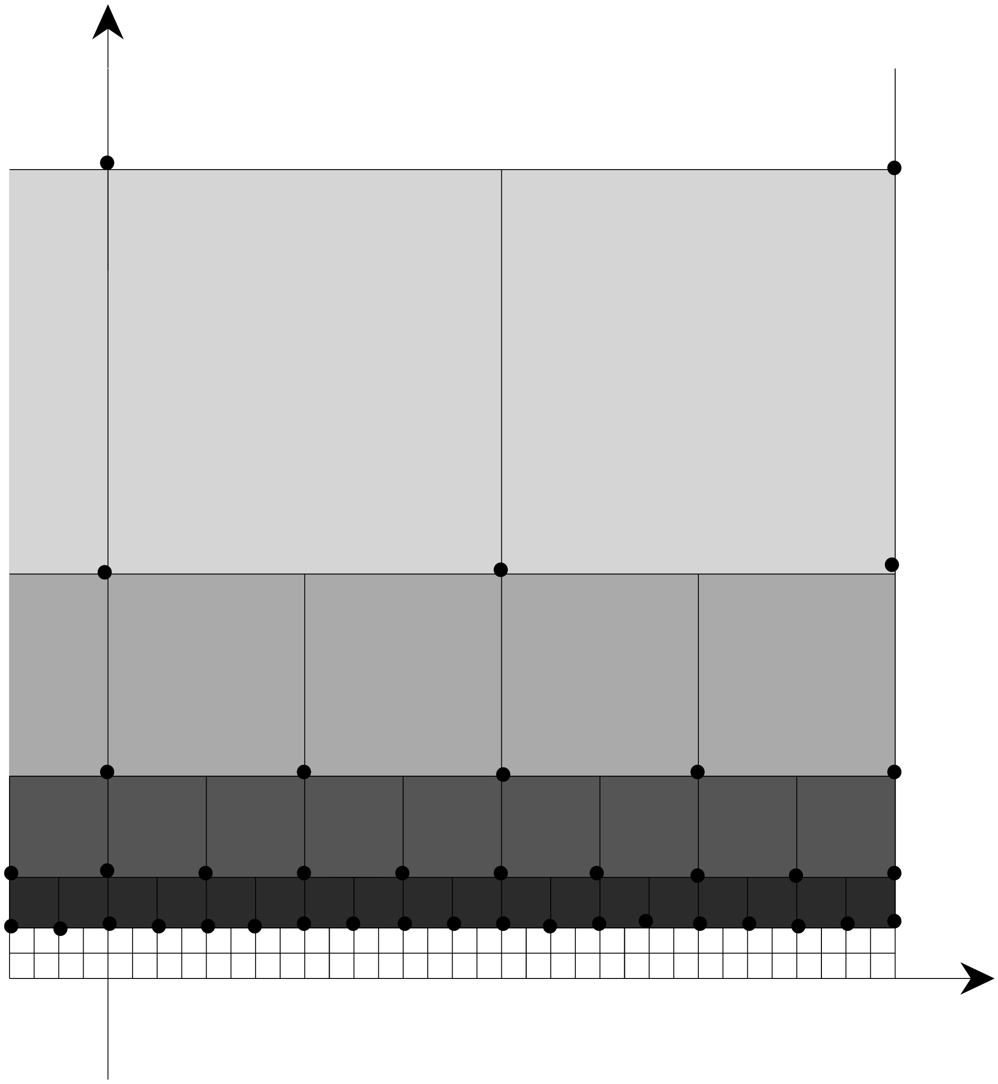


fig. 1 dyadic-plane

In figure 1 these points are the solid circles $\delta = (k/2^m,\, 1/2^m)$. This associates a *lattice* D of points δ in the upper half-plane with the set of dyadic numbers, and the basic dyadic interval $I_\delta = [k/2^m,\, (k+1)/2^m)$ is the projection onto the real axis of the *dyadic square* Δ_δ whose lower left-hand corner is the point δ. These dyadic squares are the shaded squares in figure 1. The geometric property of basic dyadic intervals says that exactly one of

$$I_\delta \cap I_{\delta'} = \phi\ , \quad I_\delta \subset I_{\delta'}\ , \quad I_\delta \supset I_{\delta'} \tag{1.43}$$

holds for any pair of different elements δ, δ' in D.

(1.44) DEFINITION. *The Haar functions h_δ are defined by*

$$h_\delta(x) = 2^{m/2} h(2^m x - k)$$

for each $\delta = (k/2^m,\ 1/2^m)$ in D.

As $h(x) = 0$ except when $0 \le x < 1$, $h_\delta(x) = 0$ except when $0 \le 2^m x - k < 1$, *i.e.*, except when $k/2^m \le x < (k+1)/2^m$. Thus the only non-zero part of the graph of $y = h_\delta$ lies above or below the basic dyadic interval I_δ, and it is clear from this that each h_δ has average zero in the sense that

$$\int_{-\infty}^{\infty} h_\delta(x)\, dx = 0 \qquad (\delta \in D)\ . \tag{1.45}$$

Thus we can think of h_δ as being the 'wave-like' function generated from the single wave-like function $h = h(x)$, the so-called *Mother Wavelet*, by *dilation* $x \to 2^m x$ and *translation* $x \to x - k$.

(1.46) THEOREM. *The Haar functions satisfy*

$$\int_{-\infty}^{\infty} h_\delta(x) h_{\delta'}(x)\, dx = \begin{cases} 0 & (\delta \ne \delta')\ , \\ 1 & (\delta = \delta')\ ; \end{cases}$$

in particular, $\{h_\delta : \delta \in D\}$ is an orthonormal family in $L^2(-\infty, \infty)$.

Proof. If $\delta = \delta'$, then

$$\int_{-\infty}^{\infty} h_\delta(x) h_{\delta'}(x)\, dx = 2^m \int_{-\infty}^{\infty} \big(h(2^m x - k)\big)^2\, dx\ .$$

Now change variables $x \to 2^m x$. Then

$$\int_{-\infty}^{\infty} h_\delta(x) h_{\delta'}(x)\, dx = \int_{-\infty}^{\infty} \big(h(x-k)\big)^2\, dx = \int_{k}^{k+1} \big(h(x-k)\big)^2\, dx = \int_{k}^{k+1} dx = 1\ .$$

On the other hand, if $\delta \ne \delta'$, then

$$\int_{-\infty}^{\infty} h_\delta(x) h_{\delta'}(x)\, dx = 0 \tag{1.47}$$

if the basic dyadic intervals $I_\delta, I_{\delta'}$ on which $h_\delta, h_{\delta'}$ live are disjoint. But, if $I_\delta \subseteq I_{\delta'}$ say, then $h_{\delta'}$ will have constant sign on I_δ; in this case

$$\int_{-\infty}^{\infty} h_\delta(x) h_{\delta'}(x)\, dx = \pm \int_{I_\delta} h_\delta(x)\, dx = 0$$

since h_δ has average 0. By the geometric property of basic dyadic intervals the only other possibility is $I_\delta \supseteq I_{\delta'}$ and (1.46) again follows because $h_{\delta'}$ has average 0. This establishes the orthonormality of the h_δ, $\delta \in D$. ∎

The orthonormal series associated with the Haar functions is given by

(1.48)
$$H[f] = \sum_{\delta \in D} (f, h_\delta) h_\delta(x) = \sum_{\delta \in D} a_\delta h_\delta(x)$$

where the Haar coefficients $\{a_\delta : \delta \in D\}$ are given by

(1.49)
$$a_\delta = \int_{I_\delta} f(t) h_\delta(t)\, dt = 2^{m/2} \int_{-\infty}^{\infty} f(t) h(2^m t - n)\, dt\ .$$

Consequently,

(1.50)
$$H[f] = \sum_{m,n} 2^m \left(\int_{-\infty}^{\infty} f(t) h(2^m t - n)\, dt \right) h(2^m x - n)\ .$$

(1.51) THEOREM. *The set of Haar functions $\{h_\delta : \delta \in D\}$ is a complete orthonormal family in $L^2(-\infty, \infty)$; in particular,*

$$f = \sum_{\delta \in D} a_\delta h_\delta\ , \qquad \int_{-\infty}^{\infty} |f(x)|^2\, dx = \sum_{\delta \in D} |a_\delta|^2\ ,$$

hold for all f in $L^2(-\infty, \infty)$.

It can be shown that

$$f(x) = \sum_{\delta \in D} a_\delta h_\delta(x) \qquad (x \in \mathbb{R})$$

holds pointwise whenever f is a continuous function; otherwise, the convergence of the infinite series $\sum_\delta a_\delta h_\delta$ has to be taken "in mean." These results illustrate the basic point of wavelet expansions, but also bring out the deficiency of the Haar wavelets $\{h_\delta : \delta \in D\}$.

If we think of a function $f = f(x)$ in the Haar series $\sum_\delta a_\delta h_\delta$ as the recovery of the wavelet expansion $f = \sum_\delta a_\delta h_\delta$ decomposes the signal into the sum of wave-like signals — single square waves each one generated by dilation and translation from a single mother wavelet h. More crucially, the coefficient a_δ depends only on the behaviour of the signal on the interval I_δ. If f is essentially constant on I_δ, then by (1.49) the coefficient a_δ will be extremely small because h_δ has average zero; on the other hand, if f oscillates on I_δ, the oscillation of h_δ on I_δ will ensure that the integrand $f(x)h_\delta(x)$ has constant sign on I_δ, and so give a large positive or negative value to a_δ. Thus the Haar coefficients can be used to detect singularities of the signal where it is oscillating very rapidly. The length 2^{-m} of I_δ is a measure of the scale of resolution of the signal: the larger m is, the higher the resolution. Technically, the Haar coefficients $\{a_\delta : \delta \in D\}$ provide a *multi-scale resolution* of the signal f that can be used to detect singularities in the signal. Mathematically, the complete orthonormality property of the Haar functions ensures that the function can be recovered exactly from the sum $\sum_\delta a_\delta h_\delta$ over *all* Haar coefficients a_δ. But we have seen that the a_δ will be insignificant over intervals I_δ on which the signal changes little. Consequently, all the essential features of the signal will still be recovered if these a_δ are omitted from the sum. Hence, for all practical purposes, the signal can be recovered from a subset of the Haar coefficients $\{a_\delta : \delta \in D\}$. If we think of the mapping $f \to \{a_\delta : \delta \in D\}$ as being a *Haar Scanner* which associates a discrete data set to each signal, and the Haar series $\sum_\delta a_\delta h_\delta$ as the recovery of the signal from this data set, *data compression* will be achieved by retaining only those a_δ of sufficient size without losing the essential features of the signal. The deficiency of such a scanner is that a possibly continuous signal is decomposed into discontinuous wavelets h_δ. A replacement for the Haar mother is needed in which the mother wavelet is smoother and yet vanishes outside some bounded set so that the wavelets can still be used to detect such *local* behaviour of signals as singularities.

The simplest continuous substitute for the Haar functions is the 'hat' function

$$\beta(x) = \begin{cases} 1 - |x| & (|x| < 1)\ , \\ 0 & (|x| \geq 1)\ , \end{cases} \tag{1.52}$$

and the translates $\{\beta(x-k) : k \in \mathbb{Z}\}$. As each $\beta(x-k)$ overlaps with $\beta(x-k+1)$ and $\beta(x-k-1)$, this family cannot be orthonormal, but it is a stable family. We leave as an exercise the proof of the following result.

(1.53) THEOREM. *When β is the 'hat' function* (1.52)*, the family* $\{\beta(x-k) : k \in \mathbb{Z}\}$ *of its integer translates is a stable family in* $L^2(\mathbb{R})$*. More precisely, the inequalities*

$$\tfrac{1}{3}\sum_k |a_k|^2 \le \int_{-\infty}^{\infty} \Big| \sum_k a_k \beta(x-k) \Big|^2 \, dx \le \sum_k |a_k|^2 \; .$$

hold for all $\{a_k\}$ *in* ℓ^2.

2. Operator theory. Just as a Hilbert space is the natural abstraction of Euclidean space so a bounded linear operator $T : \mathcal{H}_1 \to \mathcal{H}_2$ between two Hilbert spaces is the natural abstraction of linear transformations on Euclidean space, i.e., of matrices.

(2.1) DEFINITION. *A linear operator $T : \mathcal{H}_1 \to \mathcal{H}_2$ from one Hilbert space $\mathcal{H}_1$ into a possibly different Hilbert space $\mathcal{H}_2$ is said to be bounded, and have operator norm $\|T\|$, when*

$$\|T\| = \sup\{\|Tf\|_{\mathcal{H}_2} : \|f\|_{\mathcal{H}_1} \le 1\}$$

is finite. The set of all such operators will be denoted by $\mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$ or by $\mathcal{L}(\mathcal{H})$ when $\mathcal{H} = \mathcal{H}_1 = \mathcal{H}_2$.

Since this definition of bounded operator depends only on the norm on a Hilbert space, not on the inner product, it carries over immediately to linear operators $T : \mathfrak{X} \to \mathfrak{Y}$ between arbitrary Banach spaces: thus

$$\text{(2.1)(i)} \qquad \|T\| = \{\sup \|Tx\|_{\mathfrak{Y}} : x \in \mathfrak{X}\ ,\ \|x\|_{\mathfrak{X}} \le 1\}$$

and we denote by $\mathcal{L}(\mathfrak{X}, \mathfrak{Y})$ the Banach space of all such operators T for which $\|T\|$ is finite. When $\mathfrak{X} = \mathfrak{Y}$, the notation is usually abbreviated to $\mathcal{L}(\mathfrak{X})$.

Each matrix $A = [a_{ij}]$ in $\mathbb{C}^{m\times n}$ defines a linear operator $A : \mathbb{C}^m \to \mathbb{C}^n$ by matrix multiplication (on the right):

$$\text{(2.2)} \qquad A : z = (z_1, \ldots, z_m) \to (w_1, \ldots, w_n) \qquad \left(w_j = \sum_{i=1}^{m} a_{ij} z_i\right) .$$

For example the Discrete Fourier transform

$$\text{(2.3)(i)} \qquad \{z_j\} \to \{\hat{z}_k\}\ , \qquad \hat{z}_k = \frac{1}{\sqrt{n}} \sum_j z_j \bar{\omega}^{jk}\ ,$$

corresponds to matrix multiplication on the right by the $n \times n$ matrix

$$\text{(2.3)(ii)} \qquad U_D = \frac{1}{\sqrt{n}} [\bar{\omega}^{jk}]_{j,k=0}^{n-1} = \frac{1}{\sqrt{n}} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & \bar{\omega} & \cdots & \bar{\omega}^{n-1} \\ 1 & \bar{\omega}^2 & \cdots & \bar{\omega}^{2(n-1)} \\ \vdots & \vdots & & \vdots \\ 1 & \bar{\omega}^{n-1} & \cdots & \bar{\omega}^{(n-1)(n-1)} \end{bmatrix} ,$$

while the Inverse Discrete Fourier transform

$$\text{(2.3)(iii)} \qquad \{\hat{z}_k\} \to \{z_j\}\ , \quad z_j = \frac{1}{\sqrt{n}} \sum_k \hat{z}_k \omega^{jk}$$

corresponds to matrix multiplication on the right by

$$\text{(2.3)(iv)} \qquad U_D^{-1} = \frac{1}{\sqrt{n}} [\omega^{jk}]_{j,k=0}^{n-1} = \frac{1}{\sqrt{n}} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & \omega & \cdots & \omega^{n-1} \\ 1 & \omega^2 & \cdots & \omega^{2(n-1)} \\ \vdots & \vdots & & \vdots \\ 1 & \omega^{n-1} & \cdots & \omega^{(n-1)(n-1)} \end{bmatrix} .$$

There are many estimates for the operator norm $\|A\|$ of the general operator (2.2): for by the Cauchy-Schwarz inequality,

$$\begin{aligned} \|A(z)\| &= \left(\sum_{j=1}^{n} \left| \sum_{i=1}^{m} a_{ij} z_i \right|^2 \right)^{1/2} \\ &\le \left(\sum_{j=1}^{n} \left(\sum_{i=1}^{m} |a_{ij}|^2 \right) \left(\sum_{i=1}^{m} |z_i|^2 \right) \right)^{1/2} = \left(\sum_{i,j} |a_{ij}|^2 \right)^{1/2} \|z\| \ . \end{aligned}$$

Thus

$$\|A\| \le \left(\sum_{i,j} |a_{ij}|^2 \right)^{1/2} .$$

There is a large class of corresponding operators — so-called integral kernel operators — on spaces of functions. For instance, suppose $K = K(x, y)$ is a complex-valued function on $\mathbb{R} \times \mathbb{R}$ and define the associated integral kernel operator on functions on $(-\infty, \infty)$ by

$$\text{(2.4)} \qquad \mathcal{K} f(x) = \int_{-\infty}^{\infty} K(x, y) f(y)\, dy \ .$$

(Strictly speaking, this definition might better be thought of as corresponding to matrix multiplication on the left.) Repeating the Cauchy-Schwarz inequality proof for the matrix operator A we obtain the first estimate for $\|\mathcal{K}\|$.

(2.5) THEOREM. *The integral kernel operator given by (2.4) defines a bounded linear operator on $L^2(\mathbb{R})$ when*

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |K(x,y)|^2\,dx\,dy \tag{2.5)(i}$$

is finite; and when this is finite,

$$\|\mathcal{K}\| \le \left(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |K(x,y)|^2\,dx\,dy\right)^{1/2} . \tag{2.5)(ii}$$

A slightly different use of the Cauchy-Schwarz inequality gives an alternative estimate for $\|\mathcal{K}\|$.

(2.6) THEOREM. *The integral kernel operator given by (2.4) defines a bounded linear operator on $L^2(\mathbb{R})$ whenever both of*

$$\sup_x \int_{-\infty}^{\infty} |K(x,y)|\,dy \quad , \quad \sup_y \int_{-\infty}^{\infty} |K(x,y)|\,dx \tag{2.6)(i}$$

are finite; and when these are finite,

$$\|\mathcal{K}\| \le \left(\sup_x \int_{-\infty}^{\infty} |K(x,y)|\,dy\right)^{1/2} \left(\sup_y \int_{-\infty}^{\infty} |K(x,y)|\,dx\right)^{1/2} \tag{2.6)(ii}$$

Proof. By definition,

$$\begin{aligned}
\|\mathcal{K}f\|^2 &= \int_{-\infty}^{\infty} \Big|\int_{-\infty}^{\infty} K(x,y)f(y)\,dy\Big|^2\,dx \\
&\le \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} |K(x,y)|^{1/2}|K(x,y)|^{1/2}|f(y)|\,dy\right)^2 dx\ .
\end{aligned}$$

But

$$\begin{aligned}
&\int_{-\infty}^{\infty} |K(x,y)|^{1/2}|K(x,y)|^{1/2}|f(y)|\,dy \\
&\qquad \le \left(\int_{-\infty}^{\infty} |K(x,y)|\,dy\right)^{1/2}\left(\int_{-\infty}^{\infty} |K(x,y)|\ |f(y)|^2\,dy\right)^{1/2} ,
\end{aligned}$$

by the Cauchy-Schwarz inequality. Consequently,

$$\begin{aligned}
\|\mathcal{K}f\|^2 &\le \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} |K(x,y)|\,dy\right)\left(\int_{-\infty}^{\infty} |K(x,y)|\ |f(y)|^2\,dy\right) dx \\
&\le \left(\sup_x \int_{-\infty}^{\infty} |K(x,y)|\,dy\right)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |K(x,y)|\ |f(y)|^2\,dy\,dx\ ,
\end{aligned}$$

and so after changing the order of integration on the right hand side we deduce that

$$\|\mathcal{K}f\|^2 \le \left(\sup_x \int_{-\infty}^{\infty} |K(x,y)|\,dy\right) \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} |K(x,y)|\,dx\right) |f(y)|^2\,dy$$

$$\le \left(\sup_x \int_{-\infty}^{\infty} |K(x,y)|\,dy\right)\left(\sup_y \int_{-\infty}^{\infty} |K(x,y)|\,dx\right) \|f\|^2 .$$

This establishes (2.4)(ii). ∎

If Hölder's inequality is used instead of the Cauchy-Schwarz inequality, the corresponding estimate for the boundedness of $\mathcal{K}$ on $L^p(\mathbb{R})$ follows:

(2.7) $$\|\mathcal{K}\| \le \left(\sup_x \int_{-\infty}^{\infty} |K(x,y)|\,dy\right)^{1/q} \left(\sup_y \int_{-\infty}^{\infty} |K(x,y)|\,dx\right)^{1/p} .$$

So far none of these estimates has used any property of the domain of the functions in the integrand. Indeed, if (X,μ) and (Y,ν) are arbitrary measure spaces and $K = K(x,y)$ is a (measurable) function on $X \times Y$, then the integral operator

(2.8)(i) $$\mathcal{K}f(x) = \int_Y K(x,y)f(y)\,d\nu(y)$$

can be defined completely analogously to (2.4). Furthermore, the Cauchy-Schwarz or Hölder inequality proofs for (2.5), (2.6) and (2.7) carry over to this more general setting, establishing all the corresponding results:

(2.8)(ii) *as an operator* $\mathcal{K} : L^2(Y,\nu) \to L^2(X,\mu)$

(a) $$\|\mathcal{K}\| \le \left(\int_X \int_Y |K(x,y)|^2\,d\nu\,d\mu\right)^{1/2} ,$$

(b) $$\|\mathcal{K}\| \le \left(\sup_x \int_Y |K(x,y)|\,d\nu\right)^{1/2} \left(\sup_y \int_X |K(x,y)|\,d\nu\right)^{1/2} ;$$

(2.8)(iii) *as an operator* $\mathcal{K} : L^p(Y,\nu) \to L^p(X,\mu)$

$$\|\mathcal{K}\| \le \left(\sup_x \int_Y |K(x,y)|\,d\nu\right)^{1/q} \left(\sup_y \int_X |K(x,y)|\,d\mu\right)^{1/p} .$$

To illustrate these ideas recall first the *convolution*

(2.9) $$(f * g)(x) = \int_{-\infty}^{\infty} f(x-y)g(y)\,dy$$

of functions f and g on $\mathbb{R}$. We shall apply (2.5) and (2.6) with $K(x,y) = f(x-y)$. For then

$$\int_{-\infty}^{\infty} |K(x,y)|\,dy = \int_{-\infty}^{\infty} |f(x-y)|\,dy = \int_{-\infty}^{\infty} |f(y)|\,dy\ ,$$

while

$$\int_{-\infty}^{\infty} |K(x,y)|\,dx = \int_{-\infty}^{\infty} |f(x-y)|\,dx = \int_{-\infty}^{\infty} |f(x)|\,dx\ .$$

Consequently,

$$\|f * g\|_{L^p(\mathbb{R})} \le \|f\|_{L^1(\mathbb{R})} \|g\|_{L^p(\mathbb{R})}\ ; \tag{2.10}$$

in particular, $L^1(\mathbb{R})$ is a Banach algebra under convolution and $L^p(\mathbb{R})$ a Banach $L^1(\mathbb{R})$-module under convolution.

Now definition (2.9) relies on the fact that *subtraction* is a well-defined operation on $\mathbb{R}$ and, crucially, that the *translation operator* $\tau(y) : f(x) \to f(x-y)$ has norm 1 as an operator on $L^1(\mathbb{R})$. This is the first time that the structural properties of the domain of f have played a role (but it will be far from the last). Indeed, the proof of (2.10) carries over to the set $\mathbb{Z}$ of all integers, replacing (2.9) by the convolution

$$(f * g)_n = \sum_{m=-\infty}^{\infty} f_m g_{n-m} \qquad (n \in \mathbb{Z}) \tag{2.11}$$

of sequences $\{f_m\}$ and $\{g_m\}$. Again, the translation operator $\tau(m) : \{f_n\} \to \{f_{n-m}\}$ has norm 1 as an operator on $\ell^1(\mathbb{Z})$, and so

$$\|f * g\|_{\ell^p(\mathbb{Z})} \le \|f\|_{\ell^1(\mathbb{Z})} \|g\|_{\ell^p(\mathbb{Z})}\ ; \tag{2.12}$$

in particular, $\ell^1(\mathbb{Z})$ is a Banach algebra under convolution and $\ell^p(\mathbb{Z})$ is a Banach $\ell^1(\mathbb{Z})$-module under convolution.

In a seemingly different direction there is an interesting integral operator on functions defined on $(0,\infty)$:

$$(Hf)(x) = \int_0^{\infty} \frac{f(y)}{x+y}\,dy \qquad \big(f \in L^p(0,\infty)\big)\ .$$

often called Hilbert's operator though it is not to be confused with the even more subtle Hilbert transform operator. Here $K(x,y) = 1/(x+y)$, and, as simple estimates show, none of

$$\int_0^\infty \int_0^\infty |K(x,y)|^2 \, dx \, dy \quad , \quad \sup_x \int_0^\infty |K(x,y)| \, dy \quad , \quad \sup_y \int_0^\infty |K(x,y)| \, dx$$

is finite, so neither theorem (2.5) nor (2.6) applies. Nonetheless it can be shown that

$$\int_0^\infty |Hf(x)|^p \, dx \le \frac{\pi}{\sin(\frac{\pi}{p})} \int_0^\infty |f(x)|^p \, dx \qquad (1 < p < \infty) \; ;$$

in particular, H is bounded on $L^2(0,\infty)$ with norm at most $\pi^{1/2}$ (actually $\|H\| = \pi^{1/2}$).

Many of the usual definitions and results for matrices carry over to bounded linear operators on Hilbert space. For instance, the *adjoint* of a matrix $A = [a_{ij}]$ in $\mathbb{C}^{m\times n}$ is the matrix $A^* = [\bar{a}_{ji}]$ in $\mathbb{C}^{n\times m}$. In operator-theoretic terms this means that

$$\big(A(z), w\big) = \sum_j \Big(\sum_i z_i a_{ij}\Big) \bar{w}_j = \sum_i z_i \overline{\Big(\sum_j w_j \bar{a}_{ij}\Big)} = \big(z, A^*(w)\big) \; .$$

A matrix A is said to be *self-adjoint*, or *hermitian*, when $A = A^*$, i.e., when the equality

$$\big(A(z), w\big) = \big(z, A(w)\big) \qquad (z, w \in \mathbb{C}^m)$$

holds for all z, w in $\mathbb{C}^m$; such a matrix must, of course, be square. On the other hand, a square matrix U is said to be *unitary* when

$$UU^* = U^*U = I$$

where I is the $m \times m$ identity matrix; in operator-theoretic terms, therefore, $U \in \mathbb{C}^{m\times m}$ is unitary precisely when

$$\big(U(z), U(w)\big) = (z, w) \qquad (z, w \in \mathbb{C}^m)$$

holds for all z, w in $\mathbb{C}^m$. When A is self-adjoint, it is well-known that A can be diagonalized by a unitary matrix U, i.e., $A = UDU^*$, where $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_m)$ and $\lambda_1, \ldots, \lambda_m$ are the (necessarily real) eigenvalues of A. In particular,

$$\begin{aligned} \|A(z)\|^2 &= \big(A(z), A(z)\big) = \big(UDU^*(z), UDU^*(z)\big) \\ &= \Big(D\big(U^*(z)\big), D\big(U^*(z)\big)\Big) = \sum_j \lambda_j^2 |w_j|^2 \end{aligned}$$

with $U^*(z) = (w_1, \ldots, w_m)$. Since $\|U^*(z)\| = \|z\|$, it follows that

$$\|A\| = \sup\{|\lambda_j| : j = 1, \ldots, m\} \ ,$$

thereby relating the operator norm of A to the eigenvalues of A when A is unitarily diagonalizable.

Finally, this operator-theoretic interpretation of matrix theory may help us appreciate the second of the reasons for studying complete orthonormal families in the Hilbert space. Let $\mathbf{e}_1, \ldots, \mathbf{e}_m$ be the canonical basis for $\mathbb{C}^m$, i.e., $\mathbf{e}_j = (0, \ldots, 0, 1, 0, \ldots, 0)$, the single non-zero entry being in the j^{th} place. Thus for any $\mathbf{z} = \sum_i (\mathbf{z}, \mathbf{e}_i)\mathbf{e}_i$

$$A : \mathbf{z} \to \sum_j \left(\sum_i (\mathbf{z}, \mathbf{e}_i) a_{ij} \right) \mathbf{e}_j$$

But if $A = UDU^*$ with U unitary and $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_m)$, then $\{\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_m\}$, $\boldsymbol{\xi}_i = U^*(\mathbf{e}_j)$, is an orthonormal basis for $\mathbb{C}^m$ with respect to which

$$A : \mathbf{z} = \sum_i (\mathbf{z}, \boldsymbol{\xi}_i)\, \boldsymbol{\xi}_i \to \sum_i \lambda_i (\mathbf{z}, \boldsymbol{\xi}_i)\, \boldsymbol{\xi}_i \ ,$$

i.e., A acts by *multiplication* by its eigenvalues $\lambda_1, \ldots, \lambda_m$. Thus the action of A is described much more simply by choosing a different orthonormal basis for $\mathbb{C}^m$, one particularly well-adapted to the operator A. This will continue to be true in much greater generality.

The various definitions for matrices go over with little change to general operators on Hilbert space. Recall first that to each T in $\mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$ there corresponds a unique T^* in $\mathcal{L}(\mathcal{H}_2, \mathcal{H}_1)$, *the adjoint of* T, satisfying

$$(Tf, g) = (f, T^*g) \qquad (f \in \mathcal{H}_1 \ , \ g \in \mathcal{H}_2) \ .$$

(2.13) Definition. *An operator T is said to be*

(i) self-adjoint, or hermitian, when $T = T^*$ $\qquad (T \in \mathcal{L}(\mathcal{H}))$

(ii) unitary when $TT^* = T^*T = I$ $\qquad (T \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2))$.

Thus the inequality

$$(Tf, g) = (f, Tg) \qquad (f, g \in \mathcal{H})$$

holds when T is self-adjoint, while the equality

$$(Tf, Tg) = (f, g) \qquad (f, g \in \mathcal{H}_1)$$

whenever T is unitary.

Because the matrix (2.3) implementing the Discrete Fourier transform is a unitary matrix the Discrete Fourier transform $\mathcal{F} : \ell^2(I_n) \to \ell^2(I_n)$ is an example of a unitary operator since then

$$\begin{aligned}(Tw, Tz) &= \sum_j \hat{w}_j \bar{\hat{z}}_j = \frac{1}{n}\sum_j \Big(\sum_{k,\ell} w_k \bar{z}_\ell \bar{\omega}^{jk}\omega^{j\ell}\Big) \\ &= \frac{1}{n}\sum_{k,\ell} w_k \bar{z}_\ell \Big(\sum_j \bar{\omega}^{jk}\omega^{j\ell}\Big) = \sum_k w_k \bar{z}_k = (w, z)\ .\end{aligned} \tag{2.14}$$

This is the DFT version of a fundamental result in harmonic analysis known as Parseval's Theorem.

There are two additional ways of constructing linear operators on Hilbert space both of which lie at the heart of harmonic analysis and its applications to many different areas both theoretical and applied. First we continue to see how operations on the domain of a function induce bounded linear operators on Hilbert spaces of these functions. On $\mathbb{R}$ there are two fundamental operations: addition and multiplication. They are basic to all analysis on Euclidean space. It is convenient from a conceptual point of view to combine these two operations into one family of transformations

$$g = (a, b) : x \to x \circ g = ax + b \qquad (a > 0\ ,\ b \in \mathbb{R})$$

of $\mathbb{R}$. To each such g corresponds a bounded operator

$$\pi(g) : f \to \big(\pi(g)f\big)(x) = a^{1/2} f(ax + b)$$

on $L^2(\mathbb{R})$; indeed,

$$\begin{aligned}\|\pi(g)f\|^2 &= \int_{-\infty}^{\infty} |\big(\pi(g)f\big)(x)|^2\, dx \\ &= a\int_{-\infty}^{\infty} |f(ax+b)|^2\, dx = \int_{-\infty}^{\infty} |f(x)|^2\, dx = \|f\|^2\ ,\end{aligned}$$

so $\|\pi(g)\| = 1$. Even more is true: for arbitrary f, ϕ in $L^2(\mathbb{R})$,

$$\big(\pi(g)f, \pi(g)\phi\big) = a \int_{-\infty}^{\infty} f(ax+b)\,\overline{\phi(ax+b)}\,dx = \int_{-\infty}^{\infty} f(x)\,\overline{\phi(x)}\,dx = (f, \phi) \ ,$$

i.e., $\pi(g)$ is unitary. On the other hand, the composition of $\pi(g_i)$, $i = 1, 2$, with $g_i = (a_i, b_i)$, satisfies

$$\begin{aligned} \Big(\pi(g_1)\big(\pi(g_2)f\big)\Big)(x) &= a_1^{1/2}\big(\pi(g_2)f\big)(a_1x + b_1) \\ &= (a_1a_2)^{1/2} f\big(a_2(a_1x+b_1) + b_2\big) = (a_2a_2)^{1/2} f(a_1a_2x + a_2b_1 + b_2) \ , \\ &= \big(\pi(g_1g_2)f\big)(x) \end{aligned}$$

provided we *define* the product g_1g_2 by

$$g_1g_2 = (a_1, b_1)(a_2, b_2) = (a_1a_2, a_2b_1 + b_2) \ . \tag{2.15}$$

A useful way for remembering this definition is to identify the pair (a, b) with the matrix $\begin{bmatrix} 1 & b \\ 0 & a \end{bmatrix}$ and then note that (2.15) corresponds to the usual product

$$\begin{bmatrix} 1 & b_1 \\ 0 & a_1 \end{bmatrix} \begin{bmatrix} 1 & b_2 \\ 0 & a_2 \end{bmatrix} = \begin{bmatrix} 1 & a_2b_1 + b_2 \\ 0 & a_1a_2 \end{bmatrix}$$

of matrices. Notice too that the mapping $x \to ax + b$ now corresponds to the matrix product

$$[1 \ x] \to [1 \ x] \begin{bmatrix} 1 & b \\ 0 & a \end{bmatrix} \qquad (x \in \mathbb{R}) \ .$$

The explanation for these results lies in the 'affine' geometry of the Euclidean plane. It is not hard to show that the set

$$G = \big\{(a, b) : a > 0 \ , \ b \in \mathbb{R}\big\} \tag{2.16}$$

of all $g = (a, b)$ is a group, the so-called *affine group* or '$ax + b$' group, under the multiplication defined by (2.15); the identity in G is the element $(1, 0)$ and the inverse of $g = (a, b)$ is given by

$$g^{-1} = \left(\frac{1}{a}, -\frac{b}{a}\right) \tag{2.17(i)}$$

since, clearly,

$$gg^{-1} = (a, b)\left(\frac{1}{a}, -\frac{b}{a}\right) = (1, 0) \ . \tag{2.17(ii)}$$

Because of the fundamental importance of this construction to wavelet analysis, it is worthwhile collecting together its properties as a separate theorem.

(2.18) THEOREM. *For each element (a,b) in the affine group the operator*

$$\pi(a,b)f(x) = a^{1/2} f(ax+b)$$

is a unitary operator on $L^2(\mathbb{R})$ such that

(i) $\pi(a_1,b_1)\pi(a_2,b_2) = \pi(a_1a_2, a_2b_1+b_2)$

(ii) $\pi(a,b)^{-1}f(x) = \pi(a,b)^* f(x) = a^{-1/2} f(\dfrac{x-b}{a})$

In signal analysis, the action of *translation* by b in the definition of $\pi(a,b)$ corresponds to a *shift* of b in *time* while the action of *dilation* by a corresponds to a change of *scale* by a.

As our second construction of operators on a Hilbert space $\mathcal{H}$ we introduce the idea of a *frame* which is another important concept in the theory of wavelets despite the fact that the concept of frame is weaker than that of stable family or orthonormal family. Let $\{\phi_i\}_{i\in I}$ be a family in a Hilbert space $\mathcal{H}$ such that the inequality

$$\Big\| \sum_i \alpha_i\phi_i \Big\|_{\mathcal{H}} \leq B\Big(\sum_i |\alpha_i|^2\Big)^{1/2} \tag{2.19)(i}$$

holds for all $\{\alpha_i\}$ in $\ell^2(I)$. Then

$$T : \{\alpha_i\}_{i\in I} \longrightarrow \sum_i \alpha_i\phi_i \tag{2.19)(ii}$$

defines a bounded linear operator from $\ell^2(I)$ into $\mathcal{H}$ whose bound satisfies the inequality

$$\max_i \|\phi_i\|_{\mathcal{H}} \leq \|T\| \leq B\ . \tag{2.19)(iii}$$

The adjoint of T is the bounded linear operator T^* from $\mathcal{H}$ into $\ell^2(I)$ given by

$$T^* : f \longrightarrow \big\{(f,\phi_i)\big\}_{i\in I}\ , \tag{2.20)(i}$$

and the composition $\mathcal{H} \xrightarrow{T^*} \ell^2(I) \xrightarrow{T} \mathcal{H}$ defines an operator

$$S = TT^* : f \to \sum_j (f,\phi_i)\phi_i \qquad (f\in\mathcal{H}) \tag{2.20)(ii}$$

on $\mathcal{H}$. If S were the identity operator then we would immediately have a representation

$$f = \sum_j (f,\phi_i)\phi_i \tag{2.21}$$

for every f in $\mathcal{H}$. But this is rarely the case, so we introduce the following idea.

(2.22) DEFINITION. *The family* $\{\phi_i\}_{i\in I}$ *in* $\mathcal{H}$ *is said to form a frame for* $\mathcal{H}$ *if the mapping* $T : \{\alpha_i\} \to \sum_i \alpha_i\phi_i$ *maps* $\ell^2(I)$ *onto* $\mathcal{H}$*; in particular, each* f *in* $\mathcal{H}$ *can be decomposed in the form* $f = \sum_i a_i\phi_i$ *with*

$$\|f\|_{\mathcal{H}} \le \|T\| \Big(\sum_i |a_i|^2\Big)^{1/2} .$$

In view of (1.22), every complete stable family forms a frame for $\mathcal{H}$, and in this case the decomposition $f = \sum_i a_i\phi_i$ will be unique. But for a general frame, as we shall see later, the decomposition will fail to be unique; nonetheless, this failure will have important consequences.

(2.23) THEOREM. *For a family* $\{\phi_i\}_{i\in I}$ *in a Hilbert space* $\mathcal{H}$ *satisfying* (2.19)(i) *the following assertions are equivalent:*

(i) $\{\phi_i\}_{i\in I}$ *is a frame for* $\mathcal{H}$*,*

(ii) *there exist positive constants* A, B *such that the inequalities*

$$A\Big(\sum_i |(f,\phi_i)|^2\Big)^{1/2} \le \|f\|_{\mathcal{H}} \le B\Big(\sum_i |(f,\phi_i)|^2\Big)^{1/2}$$

hold for all f *in* $\mathcal{H}$*,*

(iii) *the operator*

$$S = TT^* : f \to \sum_i (f,\phi_i)\phi_i$$

is invertible on $\mathcal{H}$*.*

Property (2.23)(ii) is the one usually adopted in defining a frame for a Hilbert space, but (2.23)(iii) best serves as a definition for a frame in an arbitrary Banach space since it contains no mention of any sequence space. Granted (2.23)(iii), choose any f in $\mathcal{H}$ and set $g = S^{-1}f$. Then

$$f = S(S^{-1}f) = Sg = \sum_i (g,\phi_i)\phi_i = \sum_i (S^{-1}f,\phi_i)\phi_i = \sum_i \big(f,(S^{-1})^*\phi_i\big)\phi_i .$$

Hence if $\{\phi_i\}$ is a frame, then every f in $\mathcal{H}$ admits a representation

$$f = \sum_i a_i\phi_i , \qquad a_i = \big(f,(S^{-1})^*\phi_i\big) , \tag{2.24}$$

but now the algorithm specifying the coefficients a_i is not as simple as in the case of an orthonormal family or a stable family. Furthermore, though it is not obvious from (2.24), there is no unique choice of the coefficients $\{a_i\}$ (cf. (2.25)(i) instead).

Proof of (2.23). (i) $\Rightarrow$ (ii) Of course, the right hand inequality in (2.23)(ii) merely expresses the boundedness of $T : \{\alpha_i\} \to \sum_i \alpha_i \phi_i$ as an operator from $\ell^2(I)$ into $\mathcal{H}$. Now let

(2.25)(i)
$$F = \Big\{\{\alpha_i\} : \{\alpha_i\} \in \ell^2(I)\ ,\ \sum_i \alpha_i \phi_i = 0\Big\}$$

be the kernel of T. Then F is a closed subspace of $\ell^2(I)$, so let

(2.25)(ii)
$$E = \Big\{\{\beta_i\} : \sum_i \beta_i \bar{\alpha}_i = 0\ ,\ \{\alpha_i\} \in F\Big\}$$

be its orthogonal complement. By (1.18) E also is a closed subspace of $\ell^2(I)$, hence a Hilbert subspace of $\ell^2(I)$. We shall prove that the restriction, $T|_E$ of T to E defines a 1–1 mapping $T|_E : E \to \mathcal{H}$ from E onto $\mathcal{H}$. General Banach space theory — the open mapping theorem — will then guarantee that the inverse mapping $(T|_E)^{-1}$ is a bounded operator from $\mathcal{H}$ back into E, i.e., the inequality

(2.26)
$$\|(T|_E)^{-1} f\|_{\ell^2(I)} \le (1/A)\|f\|_{\mathcal{H}} \qquad (f \in \mathcal{H})$$

holds for some positive constant A. To show that $T|_E$ is 1–1, suppose $(T|_E)\,\boldsymbol{\alpha} = (T|_E)\,\boldsymbol{\beta}$ for some $\boldsymbol{\alpha}, \boldsymbol{\beta}$ in E. Then

$$0 = \sum_i \alpha_i \phi_i - \sum_i \beta_i \phi_i = \sum_i (\alpha_i - \beta_i)\phi_i\ ,$$

i.e., $\boldsymbol{\alpha} - \boldsymbol{\beta}$ belongs both to F and to orthogonal complement $F^{\perp}$. Hence $\boldsymbol{\alpha} - \boldsymbol{\beta} = 0$, establishing the 1–1 property of $T|_E$. On the other hand, by definition, to each f in $\mathcal{H}$ there corresponds $\mathbf{a}$ in $\ell^2(I)$ such that $f = T\mathbf{a}$. Let $\mathbf{a} = \boldsymbol{\alpha} + \mathbf{b}$ be the orthogonal sum decomposition of $\mathbf{a}$ with $\boldsymbol{\alpha}$ in E and $\mathbf{b}$ in F. Then

$$f = T\mathbf{a} = T\,\boldsymbol{\alpha} + T\mathbf{b} = T\,\boldsymbol{\alpha} + 0\ ,$$

showing that $T|_E$ maps E onto $\mathcal{H}$.

The left hand inequality in (2.23)(ii) now follows easily. Indeed, for each f in $\mathcal{H}$,

$$\begin{aligned}\|f\|^2_{\mathcal{H}} &= (f,f) = (f, T\,\boldsymbol{\alpha}) = \sum_i \bar{\alpha}_i (f,\phi_i) \\ &\le \Big(\sum_i |(f,\phi_i)|^2\Big)^{1/2} \|\,\boldsymbol{\alpha}\,\|_E = \Big(\sum_i |(f,\phi_i)|^2\Big)^{1/2} \|(T|_E)^{-1} f\|_{\ell^2(I)}\ ,\end{aligned}$$

and so (2.26) ensures that

$$\|f\|_{\mathcal{H}} \le (1/A)\Big(\sum_i |(f,\phi_i)|^2\Big)^{1/2} .$$

(ii) $\Rightarrow$ (iii) Since

$$|(T\,\boldsymbol{\alpha}, f)| = \Big|\sum_i \alpha_i(\phi_i, f)\Big| \le \|\,\boldsymbol{\alpha}\,\|_{\ell^2}\Big(\sum_i |(f,\phi_i)|^2\Big)^{1/2} ,$$

the right hand inequality in (2.23)(ii) ensures that T is bounded with operator norm at most B; in addition, the adjoint

$$T^* : f \longrightarrow \big\{(f,\phi_i)\big\}$$

of T will be bounded from $\mathcal{H}$ into $\ell^2(I)$. But the left hand inequality in (2.23)(ii) ensures that

$$\begin{aligned}A^2\|f\|^2 &\le \sum_i |(f,\phi_i)|^2 = (T^*f, T^*f) \\ &= (TT^*f, f) \le \|TT^*f\|\ \|f\| .\end{aligned}$$

Consequently, (2.23)(ii) ensures that

$$A^2\|f\|_{\mathcal{H}} \le \|TT^*f\|_{\mathcal{H}} \le B^2\|f\|_{\mathcal{H}} ,$$

i.e., TT^* is a positive bounded operator on $\mathcal{H}$. By the spectral theorem therefore, the inverse $(TT^*)^{-1}$ of TT^* also will be a bounded operator on $\mathcal{H}$.

(iii) $\Rightarrow$ (i) Given f in $\mathcal{H}$, set $\alpha = T^*(TT^*)^{-1}f$. Then

$$T\alpha = (TT^*)(TT^*)^{-1}f = f\ ,$$

completing the proof. ■

A simple finite-dimensional example brings out clearly the distinction between a stable family and a frame. Set $\mathcal{H} = \mathbb{C}^2$ and choose vectors

$$\phi_1 = (1,0)\ , \quad \phi_2(1,1)\ , \quad \phi_3 = (0,1)$$

in $\mathcal{H}$. Then

$$T : (\alpha_1, \alpha_2, \alpha_3) \longrightarrow (\alpha_1 + \alpha_2\ ,\ \alpha_2 + \alpha_3)\ ,$$

so

$$\|T(\alpha_1, \alpha_2, \alpha_3)\|_{\ell^2} \le \sqrt{6}(\alpha_1^2 + \alpha_2^2 + \alpha_3^2)^{1/2}\ ,$$

and $T : (\alpha, -\alpha, \alpha) \to 0$. Thus T is bounded as an operator from ℓ_3^2 into ℓ_2^2. But $\{\phi_1, \phi_2, \phi_3\}$ is not stable since

$$\|T(\alpha, -\alpha, \alpha)\|_{\ell^2} = \|\alpha\phi_1 - \alpha\phi_2 + \alpha\phi_3\|_{\ell^2} = 0\ .$$

On the other hand, for any $f = (\alpha, \beta)$ in $\mathcal{H}$,

$$\big\{(f, \phi_i)\big\}_{i=1}^3 = (\alpha, \alpha + \beta, \beta)\ .$$

Thus

$$(|\alpha|^2 + |\beta|^2)^{1/2} \le \left(\sum_{i=1}^3 |(f, \phi_i)|^2\right)^{1/2} \le \sqrt{3}(|\alpha|^2 + |\beta|^2)^{1/2}$$

i.e.,

$$\|f\|_{\mathcal{H}} \le \left(\sum_{i=1}^3 |(f, \phi_i)|^2\right)^{1/2} \le \sqrt{3}\|f\|_{\mathcal{H}}\ ,$$

so $\{\phi_1, \phi_2, \phi_3\}$ is a frame for $\mathcal{H}$.

Later, in section 2 of Part II, the family

(2.27) $$\big\{a^{\frac{1}{2}m}\psi(a^m x + bk) : k, m \in \mathbb{Z}\big\} \qquad (a > 1,\ b > 0)$$

will be shown to be a frame for $L^2(\mathbb{R})$ for a large class of wavelike functions which have minimal decay and smoothness properties. The choice of $a > 1$ and $b > 0$ in (2.27) will depend on ψ however. Thus each such family will be a frame for $L^2(\mathbb{R})$ although the mother function ψ generating the family will be a wavelet mother in only the rarest of cases. In applications it is likely that these frames will be extremely useful, in part because of the wide choice of possible frames — one can try to choose the frame according to the constraints on the problem to which one wishes to apply the theory of wavelets.

Finally in this section we show how a change of orthonormal basis can be used to define important operators. Set $\mathcal{H} = \mathbb{C}^n$ and let $\{\boldsymbol{\varepsilon}_0, \ldots, \boldsymbol{\varepsilon}_{n-1}\}$ be the 'modified' canonical basis in which

$$\boldsymbol{\varepsilon}_j = (0, \ldots, 0, 1, 0, \ldots, 0) \qquad (0 \le j \le n-1) ,$$

only the $(j+1)^{st}$-coordinate being non-zero. Now let

$$\omega = \exp\left(\frac{2\pi i}{n}\right)$$

be the primitive n^{th} root of unity, and set

$$\boldsymbol{\xi}_k = \frac{1}{\sqrt{n}}(1, \omega^k, \omega^{2k}, \ldots, \omega^{(n-1)k}) \qquad (0 \le k \le n-1) .$$

Clearly $\{\boldsymbol{\varepsilon}_0, \ldots, \boldsymbol{\varepsilon}_{n-1}\}$ is an orthonormal basis for $\mathcal{H}$ so that

$$\mathbf{z} \longrightarrow (z_0, \ldots, z_{n-1}) , \qquad \mathbf{z} = \sum_{j=0}^{n-1} z_j \, \boldsymbol{\varepsilon}_j$$

identifies $\mathcal{H}$ with ℓ^2_n. On the other hand,

$$(\boldsymbol{\xi}_j, \boldsymbol{\xi}_k) = \frac{1}{n} \sum_{m=0}^{n-1} \omega^{m(j-k)} = \begin{cases} 0 & (j \ne k) , \\ 1 & (j = k) , \end{cases}$$

since

$$1 + \omega^r + \omega^{2r} + \cdots + \omega^{(n-1)r} = \frac{1 - \omega^n}{1 - \omega^r} = 0$$

for any non-zero integer r. Thus $\{\boldsymbol{\xi}_0, \ldots, \boldsymbol{\xi}_{n-1}\}$ also is an orthonormal basis for $\mathcal{H}$ and

$$\mathbf{z} \longrightarrow (\hat{z}_0, \ldots, \hat{z}_{n-1}) , \qquad \mathbf{z} = \sum_{k=0}^{n-1} \hat{z}_k \, \boldsymbol{\xi}_k$$

identifies $\mathcal{H}$ with ℓ^2_n a second time. We use these two identifications to define the Discrete Fourier transform.

(2.25) DEFINITION. *The Discrete Fourier transform* $\mathcal{F} : \ell_n^2 \to \ell_n^2$ *is defined by*

$$\mathcal{F} : f = (z_0, \ldots, z_{n-1}) \longrightarrow \hat{f} = (\hat{z}_0, \ldots, \hat{z}_{n-1})$$

where

$$\sum_{j=0}^{n-1} z_j \,\boldsymbol{\varepsilon}_j = \sum_{k=0}^{n-1} \hat{z}_k \,\boldsymbol{\xi}_k \quad .$$

Since

$$(\hat{f}, \hat{g}) = \sum_{k=0}^{n-1} \hat{z}_k \bar{\hat{w}}_k = \left(\sum_k \hat{z}_k \,\boldsymbol{\xi}_k \ , \ \sum_\ell \hat{w}_k \,\boldsymbol{\xi}_k \right)$$

$$= \left(\sum_{i=0}^{n-1} z_i \,\boldsymbol{\varepsilon}_i \ , \ \sum_{j=0}^{n-1} w_j \varepsilon_j \right) = \sum_j z_j \bar{w}_j = (f, g) \ ,$$

$\mathcal{F} : f \to \hat{f}$ is a unitary mapping from ℓ_n^2 into ℓ_n^2; in particular,

$$\|f\|_{\ell_n^2} = \|\hat{f}\|_{\ell_n^2} \qquad (f \in \ell_n^2) \ .$$

This operator can be expressed also as a matrix transformation. Indeed, by the definition of the $\boldsymbol{\xi}_k$,

$$\sum_{k=0}^{n-1} \hat{z}_k \,\boldsymbol{\xi}_k = \sum_{k=0}^{n-1} \frac{1}{\sqrt{n}} \ \hat{z}_k \left(\sum_{j=0}^{n-1} \omega^{jk} \,\boldsymbol{\varepsilon}_j \right)$$

$$= \sum_{j=0}^{n-1} \left(\frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} \hat{z}_k \omega^{jk} \right) \boldsymbol{\varepsilon}_j = \sum_{j=0}^{n-1} z_j \varepsilon_j \ ,$$

and so

$$z_j = \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} \hat{z}_k \omega^{jk} \ .$$

Conversely,

$$\hat{z}_k = \frac{1}{\sqrt{n}} \sum_{j=0}^{n-1} z_j \bar{\omega}^{jk} \ ,$$

since

$$\boldsymbol{\varepsilon}_j = \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} \bar{\omega}^{jk} \,\boldsymbol{\xi}_k \ .$$

Thus $\mathcal{F} : (z_0, \ldots, z_{n-1}) \to (\hat{z}_0, \ldots, \hat{z}_{n-1})$ corresponds to matrix multiplication on the right by

$$\frac{1}{\sqrt{n}}\left[\bar{\omega}^{jk}\right]_{j,k=0}^{n-1} = \frac{1}{\sqrt{n}}\begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & \bar{\omega} & \cdots & \bar{\omega}^{n-1} \\ 1 & \bar{\omega}^2 & \cdots & \bar{\omega}^{2(n-1)} \\ \vdots & \vdots & & \vdots \\ 1 & \bar{\omega}^{n-1} & \cdots & \bar{\omega}^{(n-1)(n-1)} \end{bmatrix}$$

while the inverse transform $\mathcal{F}^{-1} : (\hat{z}_0, \ldots, \hat{z}_{n-1}) \to (z_0, \ldots, z_{n-1})$ corresponds to matrix multiplication on the right by

$$\frac{1}{\sqrt{n}}\left[\omega^{jk}\right]_{j,k=0}^{n-1} = \frac{1}{\sqrt{n}}\begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & \omega & \cdots & \omega^{n-1} \\ 1 & \omega^2 & \cdots & \omega^{2(n-1)} \\ \vdots & \vdots & & \vdots \\ 1 & \omega^{n-1} & \cdots & \omega^{(n-1)(n-1)} \end{bmatrix}$$

3. Fourier series. Trigonometric functions, i.e., periodic wave-like functions which oscillate indefinitely without any decay, are the basic building blocks used to construct Fourier series expansion. The periodic behaviour is one feature distinguishing Fourier series expansions from wavelet expansions.

Recall that a function ϕ is said to be *periodic* and have *period* 1 if the equality $\phi(x+1) = \phi(x)$ holds for all x. For instance, each member of the family

$$\{e^{2\pi inx} : n \in \mathbb{Z}\}$$

is periodic and has period 1. This periodicity ensures that the identity

$$\int_0^1 e^{2\pi imx} e^{-2\pi inx}\, dx = \int_0^1 e^{2\pi i(m-n)x}\, dx = \Big[\frac{1}{2\pi i(m-n)} e^{2\pi i(m-n)x}\Big]_0^1 = 0 \qquad (m \neq n)$$

holds; on the other hand,

$$\int_0^1 e^{2\pi imx} e^{-2\pi inx}\, dx = \int_0^1 1\, dx = 1 \qquad (m = n)\ .$$

Thus the family $\{e^{2\pi inx} : n \in \mathbb{Z}\}$ is orthonormal in $L^2[0,1]$ and the ideas of section 1 apply.

$$(f,g) \equiv \int_0^1 f\bar{g}\, dx$$

(3.1) Definition. *For each f in $L^2[0,1]$ the coefficients*

$$\widehat{f}(n) = \int_0^1 f(y) e^{-2\pi iny}\, dy \qquad (n \in \mathbb{Z})$$

of f with respect to the orthonormal family $\{e^{2\pi inx} : n \in \mathbb{Z}\}$ are called the ***Fourier coefficients*** *of f and the corresponding orthonormal series development*

$$S[f](x) = \sum_n \widehat{f}(n) e^{2\pi inx}$$

is called the ***Fourier series*** *of f.*

Such series expansions were introduced in 1807 by Joseph Fourier in an astoundingly original paper in which he used them to solve problems in heat flow. Initially their introduction was very controversial because many eminent mathematicians found it hard to accept that an arbitrary function on $[0,1]$ could be represented by indefinitely differentiable functions which oscillate periodically. In part the controversy arose because the notion of function and theory of infinite series of functions had not been fully developed at that time. As in many other instances, Fourier analysis provided the impetus for the development of much of what is the standard mathematical repertoire today.

Let

$$S_n[f](x) = \sum_{k=-n}^{n} \hat{f}(k) e^{2\pi i k x}$$

be the "n^{th}" partial sum of the Fourier series $S[f]$ of f. Being a finite sum of period 1 functions having derivatives of all orders, each partial sum is a period 1 having derivatives of all orders. The basic fundamental question in Fourier series from the time they were first introduced was to decide on the sense in which $S_n[f](x)$ converges to $f(x)$. The sense in which $S[f]$ does indeed represent f only became clear with the development of Hilbert spaces by David Hilbert around the turn of the twentieth century and the complete answer was not given until 1964-65. The fundamental L^2-results are contained in the next theorem.

(3.2) THEOREM. *The family $\{e^{2\pi i n x} : n \in \mathbb{Z}\}$ is a complete orthonormal family in $L^2[0,1]$. In particular, the equalities*

$$\text{(i)} \qquad \int_0^1 |f(x)|^2\, dx = \sum_{-\infty}^{\infty} |\widehat{f}(n)|^2 \qquad \textit{(Plancherel)}$$

$$\text{(ii)} \qquad \int_0^1 f(x)\overline{g(x)}\, dx = \sum_{-\infty}^{\infty} \widehat{f}(n)\overline{\widehat{g}(n)} \qquad \textit{(Parseval)}$$

hold for all f, g in $L^2[0,1]$, and f can be recovered from $\{\widehat{f}(n)\}$ via the Fourier series

$$f(x) = S[f](x) = \sum_{-\infty}^{\infty} \widehat{f}(n) e^{2\pi i n x}$$

in the sense that the Fourier series $S[f]$ converges to f in mean.

Some important auxiliary results are needed for the proof of theorem (3.2). By Bessel's inequality for any orthonormal family the inequality

$$\text{(3.3)(i)} \qquad \sum_{-\infty}^{\infty} |\hat{f}(n)|^2 \le \int_0^1 |f(x)|^2\, dx \qquad \big(f \in L^2[0,1]\big)$$

always holds; in particular,

$$\text{(3.3)(ii)} \qquad \lim_{n\to\infty} \hat{f}(n) = \lim_{n\to\infty} \hat{f}(-n) = 0$$

for each f in $L^2[0,1]$. Property (3.3)(ii) is a special case of the so-called Riemann-Lebesgue lemma (the general formulation and its proof are given in the exercises for this section). Writing

$$\cos 2\pi nx = \frac{1}{2}(e^{2\pi inx} + e^{-2\pi inx}) \quad , \quad \sin 2\pi nx = \frac{1}{2i}(e^{2\pi inx} - e^{-2\pi inx}) \ ,$$

we can also express (3.3)(ii) in a useful form as: for each f in $L^2[0,1]$,

$$\text{(3.3)(iii)} \qquad \begin{aligned} \lim_{n\to\infty} \int_0^1 f(y)\cos 2\pi ny\, dy &= \lim_{n\to\infty} \frac{1}{2}\{\hat{f}(n) + \hat{f}(-n)\} = 0 \\ \lim_{n\to\infty} \int_0^1 f(y)\sin 2\pi ny\, dy &= \lim_{n\to\infty} \frac{1}{2i}\{\hat{f}(n) - \hat{f}(-n)\} = 0\ . \end{aligned}$$

The next auxiliary result is important in that it specifies conditions on a function under which its Fourier series converges pointwise rather than merely in mean as theorem (3.2) asserts. It shows that pointwise convergence depends on *smoothness*. In practice of course, pointwise convergence is more useful than convergence in mean.

(3.4) THEOREM. *Let f be continuous period 1 function having a continuous derivative. Then the Fourier series $S[f](x)$ converges pointwise for every x, $-\infty < x < \infty$; furthermore,*

$$(*) \qquad \lim_{n\to\infty} S_n[f](x) = f(x) \qquad (-\infty < x < \infty)\ .$$

Thus under the hypotheses of theorem (3.4) the Fourier series not only converges pointwise but also converges to $f(x)$ at each x.

Proof of (3.4). It is sufficient to establish $(*)$. Now

$$\text{(3.5)} \qquad S_n[f](x) = \sum_{k=-n}^{n} \hat{f}(k)e^{2\pi ikx} = \int_0^1 \underbrace{\left(\sum_{k=-n}^{n} e^{2\pi ikx}e^{-2\pi ik\theta}\right)}_{D_n(x-\theta)} f(\theta)\, d\theta\ .$$

Like many proofs in Fourier analysis, the proof of $(*)$ involves the establishing of a compact expression for this last integrand and then making careful estimates of the integrand. The compact expression in the case of (3.5) is the Dirichlet kernel:

$$D_n(\theta) = \sum_{k=-n}^{n} e^{2\pi i k\theta} = 1 + \sum_{k=1}^{n}(e^{2\pi i k\theta} + e^{-2\pi i k\theta}) = 1 + 2\sum_{k=1}^{n} \cos 2\pi k\theta \ .$$

On the other hand,

$$2\sin\pi\theta\cos 2\pi k\theta = \sin(2k+1)\pi\theta - \sin(2k-1)\pi\theta \ ,$$

and so

$$\begin{aligned} 2\sin\pi\theta\sum_{k=1}^{n}\cos 2\pi k\theta &= \sum_{k=1}^{n}\big\{\sin(2k+1)\pi\theta - \sin(2k-1)\pi\theta\big\} \\ &= \sin 3\pi\theta - \sin\pi\theta + \cdots + \sin(2n+1)\pi\theta - \sin(2n-1)\pi\theta \\ &= \sin(2n+1)\pi\theta - \sin\pi\theta \ . \end{aligned}$$

Thus

$$D_n(\theta) = 1 + \frac{\sin(2n+1)\pi\theta - \sin\pi\theta}{\sin\pi\theta} = \frac{\sin(2n+1)\pi\theta}{\sin\pi\theta} \ . \tag{3.6}$$

Hence

$$\begin{aligned} f(x) - S_n[f](x) &= f(x) - \int_0^1 D_n(x-\theta)f(\theta)\,d\theta \\ &= f(x) - \int_{x-1}^{x} D_n(\phi)f(x-\phi)\,d\phi = f(x) - \int_0^1 D_n(\phi)f(x-\phi)\,d\phi \end{aligned} \tag{3.7}$$

because of the period 1 property of the integrand. But again by the period 1 property,

$$\int_0^1 D_n(\phi)\,d\phi = \int_0^1 \Big(1 + 2\sum_{k=1}^{n}\cos 2k\pi\phi\Big)\,d\phi = 1 \ .$$

Consequently,

$$f(x) - S_n[f](x) = \int_0^1 D_n(\phi)\big\{f(x) - f(x-\phi)\big\}\,d\phi \ ,$$

and so to establish $(*)$, therefore, we have to show that

$$\lim_{n\to\infty}\int_0^1 \big\{f(x) - f(x-\theta)\big\}D_n(\theta)\,d\theta = 0$$

which, because of (3.6) is the same as showing that

$$\lim_{n\to\infty}\int_0^1 \left\{\frac{f(x)-f(x-\theta)}{\sin\pi\theta}\right\}\sin(2n+1)\pi\theta\,d\theta = 0\ .$$

The term $\{f(x)-f(x-\theta)\}$ in the integrand enables us to exploit the differentiability hypothesis on f. Indeed, by L'Hôpital's rule,

$$\lim_{\theta\to 0}\left\{\frac{f(x)-f(x-\theta)}{\sin\pi\theta}\right\} = \frac{f'(x)}{\pi}\ .$$

Thus

$$g_x(\theta) = \begin{cases} \big(f(x)-f(x-\theta)\big)/\sin\pi\theta & (\theta\neq 0)\ , \\ f'(x)/\pi & (\theta = 0)\ , \end{cases}$$

defines a continuous function of θ for each fixed x. But then

$$\begin{aligned}\int_0^1 \left\{\frac{f(x)-f(x-\theta)}{\sin\pi\theta}\right\}\sin(2n+1)\pi\theta\,d\theta &= \int_0^1 g_x(\theta)\sin(2n+1)\pi\theta\,dx \\ &= \int_0^1 \big(g_x(\theta)\cos\pi\theta\sin 2n\pi\theta + g_x(\theta)\sin\pi\theta\cos 2n\pi\theta\big)\,d\theta \\ &\longrightarrow 0 \qquad (\text{as } n\to\infty)\end{aligned}$$

applying (3.3)(iii) to the function $g_x(\theta)\cos\pi\theta$ and $g_x(\theta)\sin\pi\theta$. Hence $(*)$ holds for each x, $-\infty < x < \infty$. ∎

The proof of theorem (3.2) now follows easily.

Proof of (3.2). In view of (1.27) we have to show that

$$\lim_{n\to\infty}\int_0^1 |h(x)-S_n[h](x)|^2\,dx = 0 \tag{3.8}$$

for each h in $L^2[0,1]$. To be able to use the pointwise result of (3.4), however, we first have to replace h by a continuous period 1 function having continuous first derivative. This is where, as frequently happens, the $\mathcal{L}^2$-*density* property becomes important. Indeed, to each $\varepsilon > 0$ there corresponds a continuous period 1 function f having a continuous derivative and such that

$$\|h-f\|_{L^2[0,1]} = \left(\int_0^1 |h(x)-f(x)|^2\,dx\right)^{1/2} < \varepsilon\ .$$

In this case

$$\begin{aligned}\|h - S_n[h]\| &= \|h - f + f - S_n[f] + S_n[f] - S_n[h]\| \\ &\le \|h - f\| + \|f - S_n[f]\| + \|S_n[f - h]\| \\ &\le 2\|h - f\| + \|f - S_n[f]\| < 2\varepsilon + \|f - S_n[f]\| \ ,\end{aligned}$$

since Bessel's inequality ensures that

$$\begin{aligned}\|S_n[f - h]\|^2 &= \int_0^1 \Big| \sum_{-\infty}^{\infty} \big(\hat{f}(n) - \hat{h}(n)\big) e^{2\pi i n x} \Big|^2 \, dx \\ &= \sum_{-\infty}^{\infty} |\hat{f}(n) - \hat{h}(n)|^2 \le \|f - h\|^2 \ .\end{aligned}$$

Thus it is enough to show that

$$\lim_{n\to\infty} \int_0^1 |f(x) - S_n[f](x)|^2 \, dx = 0 \tag{3.9}$$

for any continuous period 1 function f having a continuous derivative. But by (3.4), the equality

$$f(x) = \sum_{-\infty}^{\infty} \hat{f}(k) e^{2\pi i k x} \qquad (-\infty < x < \infty)$$

holds pointwise for each such f, and so

$$\int_0^1 |f(x) - S_n[f](x)|^2 \, dx = \int_0^1 \Big| \sum_{|k|>n} \hat{f}(k) e^{2\pi i k x} \Big|^2 \, dx = \sum_{|k|>n} |\hat{f}(k)|^2 \ .$$

By Bessel's inequality yet again however,

$$\sum_{-\infty}^{\infty} |\hat{f}(k)|^2 \le \int_0^1 |f(x)|^2 \, dx \ ,$$

and so

$$\lim_{n\to\infty} \sum_{|k|>n} |\hat{f}(k)|^2 = 0 \ .$$

Hence (3.9), and hence ultimately (3.8), holds completing the proof. ∎

To illustrate these ideas consider the following functions:

(i) (Saw-tooth function – linear version). Define f on $[-\frac{1}{2}, \frac{1}{2})$ by

$$f(x) = x \qquad (-\tfrac{1}{2} \le x < \tfrac{1}{2})$$

and by $f(x+1) = f(x)$ elsewhere. Then by construction it has period 1 but it is not continuous.

After integration by parts we deduce that

$$S[f] = \frac{2}{\pi}\left\{\tfrac{1}{2}\sin 2\pi x - \tfrac{1}{4}\sin 4\pi x + \cdots\right\} = \frac{1}{\pi}\sum_{k=1}^{\infty}\frac{(-1)^{k-1}}{k}\ \sin 2k\pi x$$

(cf. exercises).

(ii) (Saw-tooth function – quadratic version). Define f on $[-\frac{1}{2}, \frac{1}{2})$ by

$$f(x) = x^2 \qquad (-\tfrac{1}{2} \le x < \tfrac{1}{2})$$

and by $f(x+1) = f(x)$ elsewhere. Then

$$S[f](x) = \frac{1}{12} + \frac{1}{\pi^2}\sum_{k=1}^{\infty}\frac{(-1)^k}{k^2}\ \cos 2\pi kx\ .$$

(iii) (Square-wave function). Define f on $[-\frac{1}{2}, \frac{1}{2})$ by

$$f(x) = \begin{cases} -1 & (-\frac{1}{2} \le x < 0)\ , \\ 1 & (0 \le x < \frac{1}{2})\ , \end{cases}$$

and by $f(x+1) = f(x)$ elsewhere. Then

$$S[f](x) = \frac{4}{\pi}\sum_{k=0}^{\infty}\frac{1}{(2k+1)}\sin(2k+1)2\pi x\ .$$

(iv) (Jagged-edge function). Define f on $[-\frac{1}{2}, \frac{1}{2})$ by

$$f(x) = 1 - 2|x| \qquad (-\tfrac{1}{2} \le x < \tfrac{1}{2})$$

and by $f(x+1) = f(x)$ elsewhere. Then

$$S[f] = \frac{1}{2} + \frac{2}{\pi^2}\sum_{k=1}^{\infty}\frac{1}{(2k-1)^2}\cos 2(2k-1)\pi x\ .$$

Note that theorem (3.4) does not apply to the saw-tooth or square-wave functions because their graphs show they are not continuous. Nor does it apply to the 'jagged edge' function for it is clear from its graph that it is continuous everywhere but fails to have a derivative at every point $\pm\frac{1}{2}, \pm\frac{3}{2}, \ldots$. Such functions are, however, continuous on $[-\frac{1}{2}, \frac{1}{2}]$ except possibly at a finite number of points; furthermore, they also have continuous derivatives on $[-\frac{1}{2}, \frac{1}{2}]$ except possibly at a finite number of points. Fortunately, with a little more care the fundamental theorem can be extended to such functions.

(3.10) DEFINITION. *A function $f = f(x)$, $a < x < b$, is said to be piecewise continuous if there exist a finite set of points $a < x_1 < \cdots < x_m < b$ such that*

(i) *f is continuous at all $x \neq x_1, x_2, \ldots, x_m$,*

(ii) *the left hand limits*

$$f_-(x_j) = \lim_{\substack{h \to 0 \\ h > 0}} f(x_j - h)$$

and right hand limits

$$f_+(x_j) = \lim_{\substack{h \to 0 \\ h > 0}} f(x_j + h)$$

exist at all points $x_1, \ldots, x_m$.

The function f is said to be piecewise smooth if both f and f' are piecewise continuous.

The saw-tooth, square-wave and jagged-edge functions, for instance, are examples of piecewise smooth functions. Simple modifications to the proof of theorem (3.4) yield its more general version.

(3.11) FIRST FUNDAMENTAL THEOREM (general version). *Let f be a piecewise smooth periodic 1 function. Then the Fourier series $S[f](x)$ converges pointwise at every x, $-\infty < x < \infty$; furthermore,*

$$\lim_{n \to \infty} S_n[f](x) = f(x)$$

if f is continuous at x, whereas

$$\lim_{n \to \infty} S_n[f](x) = \tfrac{1}{2}\big(f_-(x) + f_+(x)\big)$$

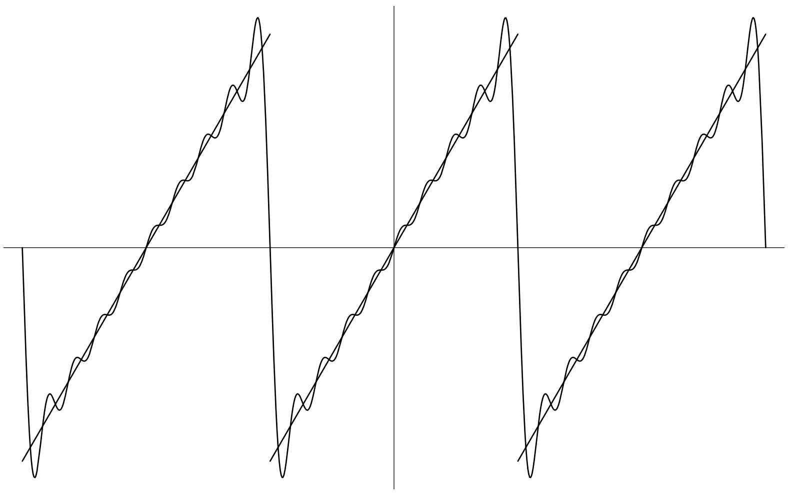
if f is not continuous at x.

In particular, therefore,

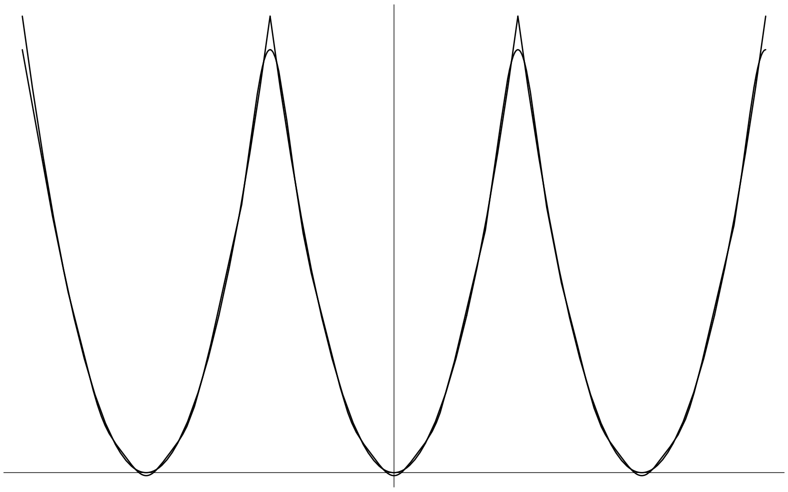
$$f(x) = \sum_{-\infty}^{\infty} \hat{f}(k) e^{2\pi i k x} \tag{3.12}$$

still holds for all x even if f is continuous but has only a piecewise continuous derivative. This result goes back to Dirichlet who proved it in 1829. It is well-illustrated by the following graphs which show how the partial Fourier sums of each function approximation the function:

(i) (Sawtooth function – linear version).



(ii) (Sawtooth function – quadratic version).



Notice that at a point of discontinuity of f the Fourier partial sums tend to overshoot the function — this is the so-called Gibbs phenomenon which was first explained thoroughly by Gibbs though it had been observed 60 years earlier by Wilbraham.

If f is continuous but fails to have a piecewise continuous derivative, then (3.12) may fail at many points. The existence of such f, first established by Du Bois-Reymond in 1876, came as a great surprise and hinted at the complexity of pointwise convergence questions for Fourier series. Indeed, the complete answer to the pointwise convergence of Fourier series was not settled until 1964-65, and not surprisingly the proof is immensely difficult.

The importance of Fourier series in applications is well-known and they are fundamental to many areas of applied mathematics wherever periodicity is present, but before turning to a study of Fourier transforms which will enable us to deal with non-periodic functions, signals or whatever the functions represent, it is worthwhile to spend a few moments putting what we have done in historical context. This in turn will help bring out one of the key technical ideas needed later. As mentioned earlier, Fourier series were introduced by Fourier at the beginning of the 19th century to solve problems in heat conduction. After much initial controversy, a great burst of activity followed in which many of the now basic concepts in mathematics were developed, in part to answer the fundamental question of the nature of the convergence of a Fourier series but also to extend the idea of representing a function as a sum or integral of special families of functions other than just the exponentials or trigonometric polynomials. For instance, Fourier series represent a function as a sum of solutions of the equation of simple harmonic motion, i.e., as eigenfunctions of $\frac{d^2}{dt^2}$, which in turn arise from separation of variables in the wave or heat equations. So it is not now surprising to find that there is a very highly developed theory of representations of functions in terms of eigenfunctions of other differential equations. Much of this development took place during the 19th century with the introduction of many families of special functions such as Bessel functions, Hermite polynomials, Legendre polynomials (actually introduced in the late 18th century) and the like. Thus the introduction of wavelet mother functions from which complete orthonormal families are generated by translation and dilation can be seen as yet another step in the process of singling out very particular functions which have an especially important role to play in mathematics. Maybe some of the wavelet mother functions (but which ones?) will soon become household words as familiar as Bessel functions and the families of special polynomials! Two other profound developments in mathematics, associated with analytic structures on the one hand and algebraic structures on the other hand, are important in understanding where we are now. The first is the introduction of Hilbert spaces by David Hilbert around 1900. We motivated their introduction as being needed for the development of a full theory of infinite orthonormal families patterned on the usual representation of vectors in 3-space by mutually orthogonal unit vectors, but Hilbert's motivation arose in studying solutions of the infinite systems of equations which came up when trying to solve integral equations connected with, say, the solution of Dirichlet's problem in differential equations by Green Function methods. Soon afterwards, Hilbert's student Hermann Weyl developed a more general theory of eigenfunction expansions using the theory of operators on Hilbert space. More recently, the most important use of Hilbert space has been in Quantum mechanics by von Neumann, Stone, Wigner, Weyl and others. Here the theory of operators on Hilbert space and especially the representation of groups by operators on particular Hilbert spaces have been the focus of much attention. Thus the mathematics in the first two sections of these notes arose from the need to develop the idea of representing a given function in terms of more special functions in a general context. The mention of group theory and of group representations brings in a more algebraic theme, seemingly. Indeed in the 1930's with the work of Pontryagin, Haar (the same one who introduced Haar functions), Weil and others Fourier theory became the study of functions on particular groups with the Fourier transform passing from functions on one group to functions on a possibly different group. In the example of Fourier series, for instance, period 1-functions can be regarded as functions on the multiplicative group while the sequences of Fourier coefficients of such functions can be regarded as functions on the additive group $\mathbb{Z}$ of all integers; by contrast, for the discrete Fourier transform the group is the same is the same in both cases - the group $\{0, 1, \ldots, n-1\}$, the group operation being the usual addition taken modulo n. In the next section we shall that for Fourier transforms the group is also the same in both cases. Functions are defined on the additive group $\mathbb{R}$ of all real numbers, while their Fourier transforms are also defined on $\mathbb{R}$. Quantum mechanics, however suggested that more complicated non-commutative groups should be considered as well, the so-called Heisenberg group being one of the most natural in this context as we shall see in the next section because of its natural connection with realizations of the Heisenberg Commutation Relations. This begins the process of returning us to the "reality" of signal analysis though, since in 1946 Gabor in a very influential paper Gabor used the Heisenberg group to introduce Time-Frequency representations of signals and the effect of the Heisenberg Uncertainty Principle A little later Ville also adopted ideas from quantum theory

4. Fourier Transforms. The theory of Fourier transforms for $\mathbb{R}$ runs parallel to the theory of Fourier series of periodic functions save for three complications:

(i) *$\mathbb{R}$ is non-compact and has infinite measure unlike $[0,1]$; in particular, $L^p(\mathbb{R})$ neither contains nor is contained in $L^1(\mathbb{R})$ for any $p > 1$;*

(ii) *the exponentials $e^{2\pi i x\xi}$ belong to $L^\infty(\mathbb{R})$ but not to any $L^p(\mathbb{R})$, $p < \infty$, and so cannot be a basis for $L^2(\mathbb{R})$,*

(iii) *a basis for $L^2(\mathbb{R})$ must necessarily be countable; i.e., must consist of some family of functions $\{\phi_i\}_{i\in I}$ with I countable.*

The complete orthonormal family of modulated bells constructed earlier provides one such countable basis — even a basis of continuous functions — while the Haar family described in section 1 provides a simpler example, though the price paid for the gain in simplicity is the lack of continuity of the basis functions. Wavelet bases will provide still further examples, but these will have (more accurately, can be constructed so that they have) all the most desirable properties of continuity, smoothness, compactness of support etc. This construction must wait upon the development of more ideas, however.

Since the $e^{2\pi i x\xi}$ do belong to $L^\infty(\mathbb{R})$, the *Fourier transform*

$$\widehat{f}(\xi) = (\mathcal{F}f)(\xi) = \int_{-\infty}^{\infty} f(x)e^{-2\pi i x\xi}\,dx \qquad (\xi \in \mathbb{R}) \tag{4.1}$$

is well-defined for f in $L^1(\mathbb{R})$, and hence for any f in $L^1(\mathbb{R})\cap L^2(\mathbb{R})$, but the definition for general f in $L^2(\mathbb{R})$ requires some care. As we shall see shortly, Plancherel's theorem for Fourier transforms states that

$$\int_{-\infty}^{\infty} |f(x)|^2\,dx = \int_{-\infty}^{\infty} |\widehat{f}(\xi)|^2\,d\xi \tag{4.2}$$

whenever f belongs both to $L^1(\mathbb{R})$ and to $L^2(\mathbb{R})$. This will enable us to extend definition (4.1) to arbitrary f in $L^2(\mathbb{R})$ by

$$\widehat{f}(\xi) = \lim_{a\to\infty} \int_{-a}^{a} f(x)e^{-2\pi i x\xi}\,dx\ , \tag{4.3}$$

the *limit being taken in mean not pointwise*. The standard proof of these results requires the development of much of the basic theory of Fourier transform theory and while this is essential for a deep understanding of the subject, Plancherel's theorem can be established in the spirit of these notes, i.e., in the context of orthonormal series. This proof also has the virtue of illustrating many of the ideas we have developed so far as well motivating the basic Fourier theory which ordinarily would precede any proof of Plancherel's theorem. Essentially, what we are doing is introducing the Short-Time Fourier transform without speaking of it as such.

(4.4) DEFINITION. *For each pair of integers* (k, m) *set*

$$b_{km}(x) = \chi(x - m)e^{2\pi ikx}$$

where χ *is the characteristic function of the interval* $[-1/2, 1/2)$.

Since the characteristic function χ can be thought of as a "discontinuous bell," the b_{km} functions have a strong similarity to the modulated bells, though the proof of the following result, the analogue of Project A in the exercises to section 3, is considerably simpler.

(4.5) THEOREM. *The functions*

$$\{b_{km} : k, m \in \mathbb{Z}\}$$

form a complete orthonormal family in $L^2(R)$.

Proof. By definition,

$$(b_{km}, b_{\ell n}) = \int_{-\infty}^{\infty} \chi(x - m)\chi(x - n)e^{2\pi i(k-\ell)x}\, dx \ .$$

Now $\chi(x - m)$ is the characteristic function of the interval $[m - 1/2, m + 1/2)$, so $\chi(x - m)$ and $\chi(x - n)$ live on disjoint sets unless $m = n$. To establish orthonormality we have only to show that

$$\int_{-\infty}^{\infty} \chi(x - m)e^{2\pi i(k-\ell)x}\, dx = \begin{cases} 1 & (k = \ell) \ , \\ 0 & (k \neq \ell) \ . \end{cases}$$

But by periodicity,

$$\int_{-\infty}^{\infty} \chi(x - m)e^{2\pi i(k-\ell)x}\, dx = \int_{0}^{1} e^{2\pi i(k-\ell)x}\, dx = \begin{cases} 1 & (k = \ell) \ , \\ 0 & (k \neq 0) \ . \end{cases}$$

This completes the proof of the $L^2(R)$-orthonormality of the family $\{b_{km}\}$.

To establish completeness we use condition (iii) in theorem (1.28) and the completeness of the exponentials $\{e^{2\pi inx} : n \in \mathbb{Z}\}$ in $L^2[0,1]$. Thus we have to show that the equality

$$\int_{-\infty}^{\infty} |f(x)|^2\, dx = \sum_{k,m} |(f, b_{km})|^2 \tag{4.6}$$

holds for every f in $L^2(R)$. But

$$(f, b_{k,m}) = \int_{-\infty}^{\infty} f(x)\chi(x-m)e^{-2\pi ik\xi} = \hat{f}_m(k)$$

writing f_m for the restriction of f to the interval $[m-1/2, m+1/2)$. Thus by theorem (3.2),

$$\int_{m-1/2}^{m+1/2} |f(x)|^2\, dx = \int_{-\infty}^{\infty} |f_m(x)|^2\, dx = \sum_{-\infty}^{\infty} |\hat{f}_m(k)|^2 \ .$$

Finally, summing over m we obtain

$$\int_{-\infty}^{\infty} |f(x)|^2\, dx = \sum_m \int_{m-1/2}^{m+1/2} |f(x)|^2\, dx = \sum_{k,m} |\hat{f}_m(k)|^2 \ .$$

This establishes (4.6) and hence the complete orthonormality of the $\{b_{km}\}$. ∎

As a consequence of (4.5), each f in $L^2(R)$ has a representation

$$f = \sum_{k,m} (f, b_{km}) b_{km} \tag{4.7}$$

convergence being taken in mean, of course. Thus, formally, the Fourier transform of f is just the linear combination

$$\hat{f}(\xi) = \sum_{k,m} (f, b_{km}) \hat{b}_{km}(\xi) \tag{4.8}$$

of the Fourier transforms of the b_{km}. The latter can be calculated explicitly:

$$\begin{aligned}
\hat{b}_{km}(\xi) &= \int_{-\infty}^{\infty} b_{km}(x) e^{-2\pi ix\xi}\, dx = \int_{-\infty}^{\infty} \chi(x-m) e^{-2\pi ix(\xi-k)}\, dx \\
&= e^{-2\pi im(\xi-k)} \int_{-\infty}^{\infty} \chi(x-m) e^{-2\pi i(x-m)(\xi-k)}\, dx \\
&= e^{-2\pi im\xi} \int_{-\infty}^{\infty} \chi(x) e^{-2\pi ix(\xi-k)}\, dx = e^{-2\pi im\xi} \hat{\chi}(\xi-k) \ .
\end{aligned}$$

On the other hand,

$$\hat{\chi}(\xi) = \int_{-1/2}^{1/2} e^{-2\pi i x\xi} = \frac{\sin \pi\xi}{\pi\xi} \ .$$

Consequently,

$$\hat{b}_{km}(\xi) = e^{-2\pi i m\xi} \, \frac{\sin \pi(\xi - k)}{\pi(\xi - k)} \ . \tag{4.9}$$

Now let us assume for the moment that the following interesting result has been proved.

(4.10) THEOREM. *The functions*

$$\{\hat{b}_{km} : k, m \in \mathbb{Z}\}$$

form an orthonormal family in $L^2(R)$*; more precisely,*

$$(\hat{b}_{km}, \hat{b}_{\ell n}) = \int_{-\infty}^{\infty} \hat{b}_{km}(\xi)\overline{\hat{b}_{\ell n}(\xi)}\, d\xi = \begin{cases} 1 & (k = \ell,\ m = n)\ , \\ 0 & \textit{(otherwise)}. \end{cases}$$

Thus the $\{b_{km}\}$ form a complete orthonormal family on the function side, so-to-speak while their Fourier transforms $\{\hat{b}_{km}\}$ form an orthonormal family on the transform or frequency side, to use signal processing language. Because of this orthonormality, however,

$$\int_{-\infty}^{\infty} \Big| \sum_{k,m} (f, b_{km})\hat{b}_{km}(\xi) \Big|^2 d\xi = \sum_{k,m} |(f, b_{km})|^2 \ . \tag{4.11}$$

Combining (4.6) and (4.11), we see that

$$\int_{-\infty}^{\infty} |f(x)|^2\, dx = \sum_{k,m} |(f, b_{km})|^2 = \int_{-\infty}^{\infty} |\hat{f}(\xi)|^2\, dx$$

for any f in $L^2(R)$. This is Plancherel's theorem for the Fourier transform. Although our proof has been somewhat formal, it is easy to make it completely rigorous by density arguments, and so establish the first part of a fundamental result in the theory.

(4.12) THEOREM. *On* $L^2(R)$ *the Fourier transform has the following properties*

(i) *(Plancherel)*

$$\int_{-\infty}^{\infty} |f(x)|^2\, dx = \int_{-\infty}^{\infty} |\hat{f}(\xi)|^2\, dx \qquad (f \in L^2)$$

(ii) *(Parseval)*

$$\int_{-\infty}^{\infty} f(x)\overline{g(x)}\,dx = \int_{-\infty}^{\infty} \hat{f}(\xi)\overline{\hat{g}(\xi)}\,dx \qquad (f, g \in L^2)\ .$$

In particular, the Fourier transform is a unitary operator from $L^2(R)$ into $L^2(R)$ in the sense that

(iii) *(unitarity)*

$$(f, g) = (\hat{f}, \hat{g}) \qquad (f, g \in L^2)\ .$$

Proof. Parseval's theorem follows immediately from Plancherel's theorem through use of the polarization identity in (1.7)(iv). The unitary property is clear also since

$$(f, g) = \sum_{k,m} (f, b_{km})\overline{(g, b_{km})} = (\hat{f}, \hat{g})\ .$$

This completes the proof. ∎

Later we shall see that the Fourier transform maps $L^2(R)$ onto $L^2(R)$ rather than merely into as theorem (4.12) asserts. There remains only the proof of theorem (4.10). We have to show that

$$(4.13) \qquad \int_{-\infty}^{\infty} \frac{\sin \pi(\xi - k)}{\pi(\xi - k)}\ \frac{\sin \pi(\xi - \ell)}{\pi(\xi - \ell)}\ e^{2\pi i(n-m)\xi}\,d\xi = \begin{cases} 1 & (k = \ell,\ m = n)\ , \\ 0 & (\text{otherwise}). \end{cases}$$

But

$$\sin \pi(\xi - k) = (-1)^k \sin \pi\xi$$

and so

$$\begin{aligned} &\int_{-\infty}^{\infty} \frac{\sin \pi(\xi - k)}{\pi(\xi - k)}\ \frac{\sin \pi(\xi - \ell)}{\pi(\xi - \ell)}\ e^{2\pi i(n-m)\xi}\,d\xi \\ &\quad = (-1)^{k-\ell} \int_{-\infty}^{\infty} \left(\frac{\sin \pi\xi}{\pi}\right)^2 \frac{1}{\xi(\xi - \ell + k)}\ e^{2\pi i(n-m)\xi}\,d\xi \\ &\quad = (-1)^{k-\ell} \int_0^1 \left(\frac{\sin \pi\xi}{\pi}\right)^2 e^{2\pi i(n-m)\xi} \left(\sum_j \frac{1}{(\xi - j)(\xi - j - \ell + k)}\right) d\xi\ . \end{aligned}$$

Suppose first that $k = \ell$. Then by the result of Project C in the exercises to section 3 the integral reduces to

$$\int_0^1 e^{2\pi i(n-m)\xi}\, d\xi$$

from which the case $k = \ell$ of (4.13) follows easily because of the orthonormality of the exponentials $\{e^{2\pi inx} : n \in \mathbb{Z}\}$ in $L^2[0,1]$. On the other hand, if $k = \ell$, then

$$\frac{1}{(\xi - j)(\xi - j - k)} = \frac{1}{k}\left\{\frac{1}{(\xi - j)} - \frac{1}{(\xi - j - k)}\right\} \qquad (k \neq 0)\ .$$

Consequently, by summing over j we see that

$$\sum_j \frac{1}{(\xi - j)(\xi - j - k)} = 0 \qquad (k \neq 0)$$

from which the remaining case $k = \ell$ of (4.13) also follows easily, completing the proof of theorem (4.10).

Now we return to the development of the more general aspects of Fourier transform theory leading up to the standard proof of Plancherel's theorem and other crucial results. Remember that to be precise we have only defined the Fourier transform for functions in $L^1(R)$. Note first that for such f its Fourier transform $\hat{f}$ is always bounded:

$$|\widehat{f}(\xi)| \leq \int_{-\infty}^{\infty} |f(x)|\, dx = \|f\|_{L^1(\mathbb{R})}\ ,$$

and, by dominated convergence,

$$\begin{aligned}\lim_{\xi\to\xi_0} \widehat{f}(\xi) &= \lim_{\xi\to\xi_0} \int_{-\infty}^{\infty} f(x) e^{-2\pi ix\xi}\, dx \\ &= \int_{-\infty}^{\infty} f(x)\left(\lim_{\xi\to\xi_0} e^{-2\pi ix\xi}\right) dx = \int_{-\infty}^{\infty} f(x) e^{-2\pi ix\xi_0}\, dx = \widehat{f}(\xi_0)\ .\end{aligned}$$

Thus $\widehat{f} = \widehat{f}(\xi)$ is a continuous function of ξ and

$$\|\widehat{f}\|_{L^\infty(\mathbb{R})} = \sup_\xi |\widehat{f}(\xi)| \leq \|f\|_{L^1(\mathbb{R})}\ . \tag{4.14}$$

Although a function and its Fourier transform are both functions on $\mathbb{R}$, it is often convenient to distinguish between the domain of a function and that of its Fourier transform. Thus we shall often write $\mathbb{R}_x$, $L^p_x(\mathbb{R}), \ldots$ on the 'function side,' so to speak, by contrast with $\mathbb{R}_\xi$, $L^p_\xi(\mathbb{R}), \ldots$ on the 'transform side.' The dual nature of $\mathbb{R}$ becomes particularly evident in the definition of the *Inverse* Fourier transform

$$\mathcal{F}^{-1}(\phi)(x) = \int_{-\infty}^{\infty} \phi(\xi) e^{2\pi i x\xi}\, d\xi \qquad (x \in \mathbb{R}) \tag{4.15}$$

which is certainly well-defined for all ϕ in $L^1_\xi(\mathbb{R})$.

Some of the basic properties of the Fourier transform are collected together in the next theorem. Notice how many of these properties already arose in the course of evaluating the Fourier transform $\hat{b}_{km}$ (cf. (4.9)).

(4.16) THEOREM. *As a linear transformation $\mathcal{F} : L^1_x(\mathbb{R}) \to \mathcal{C}(\mathbb{R})$ the Fourier transform has the properties*

(i) *(translation)*

$$\mathcal{F}\big(f(x-a)\big)(\xi) = e^{-2\pi i a\xi}(\mathcal{F}f)(\xi) \qquad (a \in \mathbb{R}) \ ,$$

(ii) *(dilation)*

$$(\mathcal{F}f_a)(\xi) = (\mathcal{F}f)(a\xi) \qquad \left(f_a(x) = \frac{1}{a} f\left(\frac{x}{a}\right)\right) \ ,$$

(iii) *(modulation)*

$$\mathcal{F}(e^{2\pi i a x} f)(\xi) = (\mathcal{F}f)(\xi - a) \qquad (a \in \mathbb{R}) \ ,$$

(iv) *(differentiation)*

$$\mathcal{F}\left(\frac{df}{dx}\right)(\xi) = 2\pi i\xi(\mathcal{F}f)(\xi) \ ,$$

(v) *(multiplication)*

$$\mathcal{F}(xf)(\xi) = \frac{-1}{2\pi i}\,\frac{d}{d\xi}(\mathcal{F}f)(\xi) \ ,$$

(vi) *(convolution)*

$$\mathcal{F}(f * g)(\xi) = (\mathcal{F}f)(\xi)(\mathcal{F}g)(\xi) \ ,$$

(vii) *(adjoint)*

$$\mathcal{F}(f^*)(\xi) = \overline{(\mathcal{F}f)\hat{}\,(\xi)} \qquad \left(f^*(x) = \overline{f(-x)}\right) \ .$$

The adjoint f^* of f gets its name from the convolution operator $T_f : g \to f * g$ on $L^2(\mathbb{R})$. For an easy calculation shows that

(4.16)(viii) $$(f * g, h) = (g, f^* * h) \qquad \big(g, h \in L^2(\mathbb{R})\big) \ ,$$

i.e., the operator T_f^* adjoint to T_f is simply T_{f^*}.

Proof. In all cases the proof is a routine exercise, as the proof of part (i) shows:

(i) after a change of variables $x \to y = x - a$,

$$\int_{-\infty}^{\infty} f(x-a) e^{-2\pi i x\xi}\, dx = e^{-2\pi i a\xi} \int_{-\infty}^{\infty} f(y) e^{-2\pi i y\xi}\, d\xi \ .$$

We leave the proof of the others as exercises. ∎

For any ϕ in $L^1(\mathbb{R})$ the dilate ϕ_t,

$$\phi_t(x) = \frac{1}{t}\phi\left(\frac{x}{t}\right) \qquad (x \in \mathbb{R} \ , \ t > 0) \ ,$$

of ϕ introduced in (4.16)(ii) plays a crucial role in Euclidean analysis. It has a number of important properties: for all t

(4.17)(i) $$\|\phi_t\|_{L^1} = \frac{1}{t}\int_{-\infty}^{\infty} |\phi\Big(\frac{x}{t}\Big)|\, dx = \int_{-\infty}^{\infty} |\phi(x)|\, dx = \|\phi\|_{L^1} \ ,$$

(4.17)(ii) $$\|f * \phi_t\|_{L^p} \le \|f\|_{L^p} \|\phi\|_{L^1} \ ,$$

(4.17)(iii) $$\widehat{\phi_t}(\xi) = \widehat{\phi}(t\xi) \qquad (\xi \in \mathbb{R}) \ .$$

Its most important property, however, is expressed in the next lemma.

(4.18) Lemma. *For each fixed ϕ in $L^1(\mathbb{R})$, the limit*

$$\lim_{t \to 0+} \|f * \phi_t - \widehat{\phi}(0) f\|_{L^p} \qquad \big(f \in L^p(\mathbb{R})\big)$$

holds for every f in $L^p(\mathbb{R})$, $1 \le p < \infty$.

Proof. Since $\widehat{\phi_t}(0) = \widehat{\phi}(0)$ for all $t > 0$,

$$\begin{aligned}(f * \phi_t)(x) - \widehat{\phi}(0) f(x) &= \int_{-\infty}^{\infty} f(x-y)\phi_t(y)\, dy - \left(\int_{-\infty}^{\infty} \phi_t(y)\, dy\right) f(x) \\ &= \int_{-\infty}^{\infty} \phi_t(y)\{f(x-y) - f(x)\}\, dy \ .\end{aligned}$$

Consequently,

$$\begin{aligned}
&|(f * \phi_t)(x) - \widehat{\phi}(0) f(x)| \\
&\qquad \le \frac{1}{t} \int_{-\infty}^{\infty} |\phi\Big(\frac{y}{t}\Big)| \; |f(x-y) - f(x)| \, dy \\
&\qquad = \int_{-\infty}^{\infty} |\phi(y)| \; |f(x-ty) - f(x)| \, dy \ ,
\end{aligned}$$

and so

$$\begin{aligned}
\|f * \phi_t - \widehat{\phi}(0) f\|_{L^p} &\le \Big\| \int_{-\infty}^{\infty} |\phi(y)| \; |f(x-ty) - f(x)| \, dy \Big\|_{L^p_x} \\
&\le \int_{-\infty}^{\infty} |\phi(y)| \; \|f(x-ty) - f(x)\|_{L^p_x} \, dy \ .
\end{aligned}$$

Now by dominated convergence

$$\lim_{t \to 0+} \int_{-\infty}^{\infty} |\phi(y)| \; \|f(x-ty) - f(x)\|_{L^p_x} \, dx = \int_{-\infty}^{\infty} |\phi(y)| \lim_{t \to 0+} \|f(x-ty) - f(x)\|_{L^p_x} \, dx \ .$$

Thus, if we can show that

$$(*) \qquad \lim_{s \to 0+} \int_{-\infty}^{\infty} |f(x-s) - f(x)|^p \, dx = 0$$

for each f in $L^p(\mathbb{R})$, $p < \infty$, then it follows immediately by dominated convergence that

$$\lim_{t \to 0+} \|f * \phi_t - \widehat{\phi}(0) f\|_{L^p} = 0 \qquad \big(f \in L^p(\mathbb{R})\big) \ .$$

Now if f were a continuous function, then the limit

$$\lim_{s \to 0} |f(x-s) - f(x)| = 0$$

would hold *pointwise*. Property $(*)$ therefore, is the L^p-substitute for functions which are not continuous but do lie in $L^p(\mathbb{R})$; thus it is usually known as the *continuity of translation property* of $L^p(\mathbb{R})$, $p < \infty$. (Its proof is left as an exercise — use the $\mathcal{L}^p(\mathbb{R})$-density property of $L^p(\mathbb{R})$.) Hence, granted property $(*)$, the lemma is established. ∎

One particular example will be useful in developing the general theory.

(4.19) LEMMA. *For each z in $\mathbb{C}$ and $c > 0$*

$$\frac{1}{(2\pi c)^{1/2}} \int_{-\infty}^{\infty} \exp\left(zx - \frac{1}{2c}x^2 \right) \, dx = \exp\left(\frac{1}{2} c z^2 \right) \; .$$

Proof. Since e^{-x^2} decays faster at infinity than e^{zx} for each fixed z in $\mathbb{C}$, the integral

$$I(z) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} \exp\left(zx - \tfrac{1}{2}x^2 \right) \, dx$$

converges for each z in $\mathbb{C}$; and for the same reason

$$\begin{aligned} \frac{d}{dz} I(z) &= \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} x \exp\left(zx - \tfrac{1}{2}x^2 \right) \, dx \\ &= \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} x e^{-\frac{1}{2}x^2} e^{zx} \, dx = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} z \exp\left(zx - \tfrac{1}{2}x^2 \right) \, dx \end{aligned}$$

after integration by parts. Thus

$$\frac{dI}{dz} = z \, I(z) \; ,$$

and so

$$I(z) = A \exp\left(\tfrac{1}{2} z^2 \right) = I(0) \exp\left(\tfrac{1}{2} z^2 \right) \; .$$

But

$$\begin{aligned} I(0)^2 &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\left(-\tfrac{1}{2}(x^2 + y^2) \right) \, dx \, dy \\ &= \frac{1}{2\pi} \int_{0}^{\infty} \int_{0}^{2\pi} \exp\left(-\tfrac{1}{2} r^2 \right) \, dr \, \theta \, dr = 1 \; . \end{aligned}$$

Hence

$$\frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} \exp\left(2x - \tfrac{1}{2}x^2 \right) \, dx = \exp\left(\tfrac{1}{2} z^2 \right) \; ,$$

from which the more general identity follows immediately by dilation $x \to x/c^{1/2}$. ∎

Setting $2\pi c = 1$ and $z = -2\pi i \xi$ in the previous lemma we deduce that

$$\int_{-\infty}^{\infty} e^{-\pi x^2} e^{-2\pi i x \xi} \, dx = e^{-\pi \xi^2} \qquad (\xi \in \mathbb{R}) \; ;$$

on the other hand, setting $2\pi c = 1$ and $z = 2\pi i s$ we see also that

$$\text{(4.20)(i)} \qquad \int_{-\infty}^{\infty} e^{-\pi \xi} e^{2\pi i s \xi} \, d\xi = e^{-\pi x^2} \; .$$

Hence the so-called Gauss-Weierstrass function

$$\sigma(x) = e^{-\pi x^2} \qquad (x \in \mathbb{R}) \tag{4.20(ii)}$$

is self-dual in the sense that

$$\sigma(x) \xrightarrow{\mathcal{F}} \sigma(\xi) \xrightarrow{\mathcal{F}^{-1}} \sigma(x) \ . \tag{4.20(iii)}$$

Many important results are a consequence of (4.18) and (4.19). Define the *Gauss-Weierstrass kernel* $W(x, y; t)$ by

$$W(x, y; t) = \sigma_{(4\pi t)^{1/2}}(x - y) = \frac{1}{(4\pi t)^{1/2}} e^{-\frac{(x-y)^2}{4t}} \ . \tag{4.21}$$

Then

$$\frac{\partial W}{\partial t} = \frac{\partial^2 W}{\partial x^2} \ ,$$

so W is a solution of the heat equation as a function of (x, t); in particular, for each f in $L^p(\mathbb{R})$

$$U(x, t) = \int_{-\infty}^{\infty} W(x, y; t) f(y) \, dy = (\sigma_{(4\pi t)^{1/2}} * f)(x)$$

also is a solution of the heat equation. But by (4.18),

$$\lim_{t \to 0+} U(x, t) = f(x)$$

provided the limit is taken in L^p (not necessarily pointwise). Hence $U = \sigma_{(4\pi t)^{1/2}} * f_{L^p(\mathbb{R})}$ is the temperature distribution at time t given an initial $L^p(\mathbb{R})$-temperature distribution $f = f(x)$. This is one physical interpretation of (4.18). To establish more theoretical consequences we first prove

(4.22) Lemma. *For each g in $L^1(\mathbb{R})$,*

$$\int_{-\infty}^{\infty} (g * \sigma_t)^\wedge(\xi) e^{2\pi i y \xi} \, d\xi = (g * \sigma_t)(y)$$

i.e., $\mathcal{F}^{-1}(\mathcal{F}(g * \sigma_t)) = g * \sigma_t$.

Proof. By definition and (4.16)(ii)

$$\begin{aligned}(g * \sigma_t)^\wedge(\xi) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(z)\sigma_t(x-z)e^{-2\pi i x\xi}\,dz\,dx \\ &= \int_{-\infty}^{\infty} g(z)\left(\int_{-\infty}^{\infty}\sigma_t(x-z)e^{-2\pi i(x-z)\xi}\,dx\right)e^{-2\pi i z\xi}\,dz \\ &= \int_{-\infty}^{\infty} g(z)\sigma(t\xi)e^{-2\pi i z\xi}\,dz\ .\end{aligned}$$

Consequently, after a change in the order of integration we see that

$$\begin{aligned}\int_{-\infty}^{\infty}(g*\sigma_t)^\wedge(\xi)e^{2\pi i y\xi}\,d\xi &= \int_{-\infty}^{\infty} g(z)\left(\int_{-\infty}^{\infty}\sigma(t\xi)e^{2\pi i(y-z)\xi}\,d\xi\right)dz \\ &= \int_{-\infty}^{\infty} g(z)\sigma_t(y-z)\,dz = (g*\sigma_t)(y)\ ,\end{aligned}$$

completing the proof. ∎

The particular case $y = 0$ in (4.22):

(4.22)(i) $$(g*\sigma_t)(0) = \int_{-\infty}^{\infty} g(x)\sigma_t(x)\,dx = \int_{-\infty}^{\infty}\widehat{g}(\xi)\widehat{\sigma}_t(\xi)\,d\xi$$

will be useful shortly. Notice already that this is already a special case of Parseval's theorem for the Fourier transform.

(4.23) THEOREM (uniqueness). *If f is an L^1-function such that $\widehat{f} \equiv 0$, then necessarily $f = 0$, i.e., $\mathcal{F}$ is 1–1.*

Proof. Suppose $\widehat{f} \equiv 0$. Then by lemma (4.12) and (4.16)(vi),

$$f*\sigma_t = \mathcal{F}^{-1}\big(\mathcal{F}(f*\sigma_t)\big) = \mathcal{F}^{-1}(\widehat{f}.\widehat{\sigma}_t) = 0\ .$$

On the other hand, $f*\sigma_t \to f$ as $t \to 0$. Thus $f = 0$. ∎

(4.24) THEOREM. *For each f in $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$,*

(4.24)(i) $$\int_{-\infty}^{\infty}|f(x)|^2\,dx = \int_{-\infty}^{\infty}|\widehat{f}(\xi)|^2\,d\xi\ ,$$

while

$$(4.24)(\text{ii}) \qquad \int_{-\infty}^{\infty} \widehat{f}(\xi)\overline{\widehat{g}(\xi)}\, d\xi = \int_{-\infty}^{\infty} f(x)\overline{g(x)}\, dx$$

holds for all f, g in $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$.

Proof. It is enough to establish (4.24)(i) since (4.24)(ii) will follow immediately by polarization. So choose any function which belongs simultaneously to $L^1(\mathbb{R})$ and $L^2(\mathbb{R})$. Then $f * \sigma_s$ also belongs to $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ and

$$(a) \qquad (f * \sigma_s)^\wedge(\xi) = \widehat{f}(\xi)e^{-\pi s^2 \xi^2} \ ,$$

while

$$(b) \qquad \int_{-\infty}^{\infty} (f * \sigma_s)(x)\overline{(f * \sigma_t)(x)}\, dx = (f * \sigma_s, f * \sigma_t)$$

$$= (f^* * f, \sigma_t * \sigma_s^*) = \int_{-\infty}^{\infty} (f * f^*)(x)\sigma_{(s^2+t^2)^{1/2}}(x)\, dx$$

because of (4.16)(viii) and the identity

$$(\sigma_t * \sigma_s^*)^\wedge(\xi) = e^{-\pi s^2 \xi^2} e^{-\pi t^2 \xi^2} = e^{-\pi(s^2+t^2)\xi^2} \ .$$

Consequently, by (4.16)(vii) and (4.22)(i),

$$\int_{-\infty}^{\infty} (f * \sigma_s)(x)\overline{(f * \sigma_t)(x)}\, dx = \int_{-\infty}^{\infty} |\widehat{f}(\xi)|^2 e^{-\pi(s^2+t^2)\xi^2}\, d\xi$$

$$= \int_{-\infty}^{\infty} (f * \sigma_s)^\wedge(\xi)\overline{(f * \sigma_t)^\wedge(\xi)}\, d\xi \ ,$$

i.e.,

$$(f * \sigma_s, f * \sigma_t) = \big(\mathcal{F}(f * \sigma_s), \mathcal{F}(f * \sigma_t)\big) \qquad (s, t > 0) \ .$$

Thus

$$\|f * \sigma_s - f * \sigma_t\|_{L^2_x} = \|\mathcal{F}(f * \sigma_s) - \mathcal{F}(f * \sigma_t)\|_{L^2_\xi} \ .$$

Hence the $L^2_x(\mathbb{R})$-version of (4.18) as well as the completeness of $L^2_\xi(\mathbb{R})$ ensures that there exists a function ψ in $L^2_\xi(\mathbb{R})$ such that

$$\lim_{t \to 0} \|\mathcal{F}(f * \sigma_t) - \psi\|_{L^2_\xi} = 0 \ , \qquad \|f\|_{L^2_x} = \|\psi\|_{L^2_\xi} \ .$$

On the other hand, by (4.18) applied to $L^1_x(\mathbb{R})$ as well as (4.14),

$$\lim_{t\to 0+} \mathcal{F}(f * \sigma_t)(\xi) = \widehat{f}(\xi)$$

uniformly in ξ. Thus $\psi = \widehat{f}$ (as an element of $L^2_\xi(\mathbb{R})$) and

$$\int_{-\infty}^{\infty} |f(x)|^2 \, dx = \int_{-\infty}^{\infty} |\widehat{f}(\xi)| \, d\xi \ ,$$

establishing (4.24)(i). ∎