# Wavelets: Math 212b, Spring 2,001

Shlomo Sternberg

May 13, 2001

# Contents

# Chapter 1

# Finite wavelets

## 1.1 Basic definitions and notations.

Let $G$ be a finite abelian group. For $z, w \in \mathcal{F}(G)$ we define the scalar product

$$\langle z, w \rangle := \sum_{a \in G} z(a)\overline{w(a)}$$

(without the normalization factor of $1/|G|$). The Fourier transform is written as

$$\hat{z}(\chi) := \langle z, \chi \rangle$$

where $\chi$ is any character on $G$. The convolution $z \star w$ of two elements of $\mathcal{F}(G)$ is defined to be

$$(z \star w)(a) = \sum_b z(ab^{-1})w(b) = \sum_c z(c)w(c^{-1}a).$$

As usual, the Fourier transform takes convolution into multiplication:

$$(z \star w)\hat{} = \hat{z}\hat{w}.$$

Indeed,

$$(z \star w)\hat{}(\chi) = \sum_{a,b} z(ab^{-1})w(b)\overline{\chi(a)} = \sum_{b,c} z(c)w(b)\overline{\chi(c)\chi(b)} = \hat{z}(\chi)\hat{w}(\chi).$$

For the group $\mathbf{Z}/N\mathbf{Z}$ where $N = 2^k$ is a power of 2 there is a famous algorithm known as the Fast Fourier Transform which computes the Fourier transform using at most $\frac{1}{2}N \log_2 N$ multiplications. We can thus compute the convolution of two functions on this group using at most $N \log_2 N + N$ multiplications. The constructions to be described below all involve convolutions, and so are *fast* from the point of view of computational complexity. In fact, the actual implemetation of the recursive algorithm as described at the end of this chapter involve $O(N)$ multiplications, as can easily be checked. So the implemented wavelt algorithm is faster than the FFT. It is "real time".

Back to general considerations: Let $R_k$ denote the operation of translation of a function by $k$, so

$$R_k z(a) := z(ak^{-1})$$

or $z(a-k)$ if we write the multiplication additively. Thus

$$(R_k z)\hat{}\,(\chi) = \langle R_k z, \chi\rangle = \langle z, R_{k^{-1}}\chi\rangle = \overline{\chi(k)}\hat{z}(\chi).$$

Define

$$\tilde{w}(a) = \overline{w(a^{-1})}. \tag{1.1}$$

Then

$$\begin{aligned}(\tilde{w})\hat{}\,(\chi) &= \langle \overline{w}(\bullet^{-1}), \chi\rangle \\ &= \langle \overline{w}, \chi(\bullet^{-1})\rangle \\ &= \langle \overline{w}, \overline{\chi}\rangle\end{aligned}$$

so

$$(\tilde{w})\hat{} = \overline{\hat{w}}. \tag{1.2}$$

Also

$$(z \star \tilde{w})(a) := \sum_b z(ab^{-1})\overline{w}(b^{-1}) = \sum_c z(c)\overline{w}(ca^{-1})$$

by making the change of variables $c := ab^{-1}$ in the middle sum. So

$$z \star \tilde{w}(a) = \langle z, R_a w\rangle \tag{1.3}$$

For this reason $z \star \tilde{w}$ is sometimes called the "correlation" of $z$ and $w$. Its value at $a$ is the scalar product of $z$ with the function $w$ translated by $a$.

Finally, $\langle \chi z, \tau\rangle = \langle z, \chi^{-1}\tau\rangle$ so

$$(\chi z)\hat{} = R_\chi \hat{z}. \tag{1.4}$$

## 1.2 The key construction.

Let $H$ be a subgroup of $G$ of index $q$. Restriction is a linear map from functions on $G$ to functions on $H$. We will denote this operator by $D$, standing for "downsampling". In the literature it is also sometimes denoted by $\downarrow$.

The restriction of a character of $G$ to $H$ is a character of $H$. Let $\mathcal{C}$ denote the kernel of this restriction map. So

$$\mathcal{C} := \ker D: \ \hat{G} \to \hat{H}$$

consists of those characters which take on the value one at all elements of $H$. We may identify $\mathcal{C}$ with the character group of $G/H$. So if

$$\Phi := \sum_{\mathcal{C}} \chi \tag{1.5}$$

then

$$\Phi(a) = \begin{cases} q & \text{if } a \in H \\ 0 & \text{if } a \notin H \end{cases} . \tag{1.6}$$

Here $q$ is the index of $H$ in $G$ which is the same as the order of the quotient group $G/H$.

We want to find $u_1, \ldots, u_q \in \mathcal{F}(G)$ such that their translates by elements of $H$ form an orthonormal basis $\mathbf{u}$ of $\mathcal{F}(G)$. Using (1.3) and the translation invariance of the scalar product, this is the requirement that

$$(u_i \star \tilde{u_j})(h) = \delta_{ij}\delta^H(h),$$

where $\delta^H$ denotes the delta function of $H$ (at the identity).

By (1.6), this amounts to

$$\Phi \cdot (u_i \star \tilde{u_j}) = q\delta_{ij}\delta^G.$$

Taking the Fourier transform and using (1.2) and (1.4) this says that

$$\sum_{\tau \in \mathcal{C}} R_\tau \hat{u}_i R_\tau \overline{\hat{u}}_j = q\delta_{ij} \tag{1.7}$$

since the Fourier transform of $\delta^G$ is the constant 1. We can rephrase this as saying that the matrix

$$A(\chi) := \frac{1}{\sqrt{q}} \begin{pmatrix} R_{\tau_1}\hat{u}_1(\chi) & \cdot & \cdot & \cdot & R_{\tau_1}\hat{u}_q(\chi) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ R_{\tau_q}\hat{u}_1(\chi) & \ldots & \cdots & \cdots & R_{\tau_q}\hat{u}_q(\chi) \end{pmatrix} \quad \tau_i \in \mathcal{C}, \tag{1.8}$$

be unitary for all $\chi$. Indeed, (1.7) says that the columns of $A(\chi)$ are orthonormal for all $\chi$.

Suppose that this condition is satisfied, so that the translates of $u_1, \ldots, u_q$ do form an orthonormal basis $\mathbf{u}$ of $\mathcal{F}(G)$. Recall that

$$D : \mathcal{F}(G) \to \mathcal{F}(H)$$

denotes restriction. Let

$$U : \mathcal{F}(H) \to \mathcal{F}(G)$$

denote extension by zero. $U$ stands for "upsampling" and is sometimes denoted in the literature by $\uparrow$. Now (1.6) says that multiplication by $\Phi$ equals $qUD$. Thus by (1.3) we have

$$UD(z \star \tilde{w})(a) = \frac{1}{q}\Phi(z \star \tilde{w})(a) = \begin{cases} \langle z, R_h w \rangle & \text{if } a = h \in H \\ 0 & \text{otherwise} \end{cases} \tag{1.9}$$

so we have from the definition of convolution,

$$z = \sum_j [UD(z \star \tilde{u_j})] \star u_j. \tag{1.10}$$

Indeed the coefficients of a $z \in \mathcal{F}(G)$ relative to the basis $\mathbf{u}$ are given by the values $\langle z, R_h u_j \rangle = z \star \tilde{u}_j(h)$ while the right hand side of (1.10) is

$$\sum_j \sum_{h \in H} (z \star \tilde{u}_j)(h) u_j(h^{-1}\bullet) = \sum_{h \in H} \sum_j \langle z, R_h u_j \rangle R_h u_j$$

which is just the expansion of $z$ in the basis $\mathbf{u}$. Thus the map

$$\mathcal{F}(G) \to \mathcal{F}(H) \oplus \cdots \oplus \mathcal{F}(H) \quad q \text{ summands}$$

$$z \mapsto D(z \star \tilde{u_1}) \oplus \cdots \oplus D(z \star \tilde{u_q})$$

gives the coefficients of $z$ relative to the $\mathbf{u}$ basis. This map is called **analysis**. The inverse map, given by (1.10) is called **synthesis**.

We emphasize again that both analysis and synthesis are given by convolutions, and hence are "fast" operations because of the Fast Fourier Transform.

## 1.3 Examples with $G = \mathbf{Z}/2M\mathbf{Z}$, and $H = 2G$.

We set $N = 2M$. Here $q = 2$ so we will write $u, v$ instead of $u_1, u_2$.

So the $A$ of (1.8) is a two by two matrix. The set $\mathcal{C}$ contains two elements: the trivial character and the character

$$n \mapsto e^{2\pi i n M/N} = e^{\pi i n}.$$

Thus

$$A(\chi_k) = \frac{1}{\sqrt{2}} \begin{pmatrix} \hat{u}(k) & \hat{v}(k) \\ \hat{u}(k+M) & \hat{v}(k+M) \end{pmatrix} \tag{1.11}$$

where $\chi_k$ denotes the character

$$\chi_k(n) = e^{2\pi i k n/N},$$

and we have written $\hat{u}(k)$ instead of $\hat{u}(\chi_k)$. It is clearly enough to check the unitarity of this matrix for $k = 0, 1, \ldots M-1$.

Also, if we have found a $u$ such that the first column of the matrix $A(\chi_k)$ is a unit vector for $k = 0, 1, \ldots, M-1$ it is easy to find a $v$ such that the matrix $A$ is unitary. Indeed, suppose we set

$$v = R_\ell[(\chi_M u)\tilde{}]$$

where $\ell$ is odd. Then

$$\hat{v}(k) = [e^{-2\pi i k \ell/N} \overline{\hat{u}}(k+M)]$$

while

$$\hat{v}(k+M) = -[e^{-2\pi i k \ell/N} \overline{\hat{u}}(k)].$$

So $\hat{v}(k)$ is a unit vector because $u(k)$ and is orthogonal to $\hat{u}(k)$.

In some of the examples below, $u$ is of the form

$$u = \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_{r-1} \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

where $r$ is even and the $a_i$ are real. Then

$$\chi_M u = \begin{pmatrix} a_0 \\ -a_1 \\ a_2 \\ \vdots \\ -a_{r-1} \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

and so

$$(\chi_M u)\tilde{} = \begin{pmatrix} a_0 \\ 0 \\ \vdots \\ -a_{r-1} \\ a_{r-2} \\ \vdots \\ a_2 \\ -a_1 \end{pmatrix}.$$

Translating by $\ell = -(r-1)$ gives

$$v = \begin{pmatrix} -a_{r-1} \\ a_{r-2} \\ \vdots \\ -a_1 \\ a_0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

So the description of $v$ is very simple: If $u$ has an even number of non-zero entries, write these in reverse order and multiply every other entry by $-1$. Of course, multiplying by an overall factor of $-1$ will not affect the outcome.

### 1.3.1 Haar wavelets.

$$u = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad v = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

The vectors $u$ and $v$ are clearly orthonormal, and translations by even integers yield pairwise orthogonal vectors. The effect of the analysis is, roughly speaking, to decompose into the trend and the variation in the "signal" $z$. The term

$$\langle z, R_h u \rangle D(R_h u)$$

produces a vector of half the length of $z$ given by

$$\frac{1}{2} \begin{pmatrix} z_0 + z_1 \\ z_2 + z_3 \\ \vdots \\ z_{N-2} + z_{N-1} \end{pmatrix}$$

while

$$\langle z, R_h v \rangle D(R_h v)$$

produces a vector of half the length of $z$ given by

$$\frac{1}{2} \begin{pmatrix} z_0 - z_1 \\ z_2 - z_3 \\ \vdots \\ z_{N-2} - z_{N-1} \end{pmatrix}.$$

### 1.3.2 Daub4

Let $a_i = a_i(\theta),\ i = 0, 1, 2, 3$ be defined by

$$\begin{aligned}
a_0 &:= \frac{1}{2\sqrt{2}} \left(1 + \sqrt{2}\cos[\theta + \frac{\pi}{4}]\right) \\
a_1 &:= \frac{1}{2\sqrt{2}} \left(1 + \sqrt{2}\cos[\theta - \frac{\pi}{4}]\right) \\
a_2 &:= \frac{1}{2\sqrt{2}} \left(1 - \sqrt{2}\cos[\theta + \frac{\pi}{4}]\right) \\
a_3 &:= \frac{1}{2\sqrt{2}} \left(1 - \sqrt{2}\cos[\theta - \frac{\pi}{4}]\right)
\end{aligned}$$

and set

$$u := \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad v := \begin{pmatrix} -a_3 \\ a_2 \\ -a_1 \\ a_0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

### 1.3.3 Shannon

Suppose that $N$ is divisible by 4. Define $u, v$ by their Fourier transforms:

$$\hat{u}(n) = \begin{cases} \sqrt{2} \text{ if } n = 0, \pm 1, \ldots, \pm(\frac{N}{4} - 1), \frac{3N}{4} \\ 0 \quad \text{otherwise} \end{cases},$$

$$\hat{v}(n) = \begin{cases} 0 \text{ if } n = 0, \pm 1, \ldots, \pm(\frac{N}{4} - 1), \frac{3N}{4} \\ \sqrt{2} \quad \text{otherwise} \end{cases}.$$

At each value of $n$ either $\hat{u}$ or $\hat{v}$ vanishes. So the columns of $A$ are orthogonal for every $n$. Also, for each $n$ either $\hat{u} = \sqrt{2}$ and $\hat{u}(n + M) = 0$ or vice versa. Thus the matrices $A(n)$ are all unitary. The effect of projection onto the space spanned by the $R_h u$ is to eliminate the "high" frequencies centered about $N/2$ while projection onto the space spanned by the $R_h v$ is to eliminate the low frequencies. Since the definitions of $\hat{u}$ and $\hat{v}$ are not conjugate symmetric about 0, the $u$ and $v$ so obtained are not real. But a slight modification in the definition will make them real.

## 1.4 Folding

Suppose $u_1$ and $v_1$ satisfy our conditions for $N$. On $H$ (corresponding to $N/2$), define

$$u_2(n) := u_1(n) + u_1(n + \frac{N}{2}), \quad v_2(n) := v_1(n) + v_1(n + \frac{N}{2}).$$

We have

$$\hat{u}_2(m) := \sum_{n=0}^{(N/2)-1} u_2(n) e^{-2\pi i n m/(N/2)} =$$

$$\sum_{n=0}^{(N/2)-1} u_1(n) e^{-2\pi i n 2m/N} + \sum_{N/2}^{N-1} u_1(n) e^{-2\pi i n 2m/N}$$

so

$$\hat{u}_2(m) = \hat{u}_1(2m) \tag{1.12}$$

with a similar result for $v$. So if $A_1(n)$ is unitary for all $n$, then $A_2(m)$ is unitary for all $m$.

## 1.5 Iteration.

We can iterate the procedure if we have a sequence of subgroups. If, $G = \mathbf{Z}/N\mathbf{Z}$ with $N = 2^p M$, for example $N = 2^p$ we can iterate $p$ times. In general there will be a tree of choices to decide which of the $u$ or $v$ at any level is decomposed further. Commonly, as in the case of the Haar wavelets, we might consider the $u$ as representing the "trend" and the $v$ as representing the variation at each stage. If (following the conventions of scattering theory) we let the index of subspaces increase with the increase of the subspace, then the original (standard or "Euclidean") space will be denoted by $A^0$, the first trend space (spanned by the translates $R_h u_1$) by $A^{-1}$ so we get a cascade:

$$
\begin{aligned}
A^0 &= A^{-1} \oplus D^{-1} \\
&= A^{-2} \oplus D^{-2} \oplus D^{-1} \\
&= \vdots \\
&= A^{-p} \oplus D^{-p} \oplus \cdots \oplus D^{-1}.
\end{aligned}
$$

The spaces $D^j$ are orthogonal, and increasing the index on the $A$ increases the detail. In terms of the corresponding basis, "compression" is achieved if many of the coefficients in terms of the constructed basis are small. Then retain only those terms corresponding to coefficients above a certain cutoff. Also, if "noise contamination" is regarded as small, this same procedure eliminates noise.

# Chapter 2

# Wavelets on $\mathbf{R}^n$

## 2.1 Multiresolution analysis

Data: $\mathbf{R}^n$ with its integer lattice $\mathbf{Z}^n$ and an invertible matrix $A$ with integer entries all of whose eigenvalues have absolute value $> 1$. Define the unitary operator

$$U_A : L_2(\mathbf{R}^n) \to L_2(\mathbf{R}^n)$$

by

$$U_A f(x) := |\det A|^{\frac{1}{2}} f(Ax).$$

Also, for $\gamma \in \mathbf{Z}^n$ define

$$T_\gamma : L_2(\mathbf{R}^n) \to L_2(\mathbf{R}^n), \quad T_\gamma f(x) := f(x - \gamma).$$

A **multiresolution analysis** associated with $A$ is a subspace

$$V_0 \subset L_2(\mathbf{R}^n)$$

such that

1. $$V_0 \subset U_A V_0.$$

   We may then define

   $$V_j := (U_A)^j V_0 \quad \forall j \in \mathbf{Z}$$

   so that

   $$\cdots \subset V_{-1} \subset V_0 \subset V_1 \subset \cdots .$$

   We then demand that

2. $\bigcup V_j$ is dense in $L_2(\mathbf{R}^n)$,
3. $\bigcap V_j = \{0\}$,
4. $T_\gamma V_0 = V_0 \ \ \forall \gamma \in \mathbf{Z}^n$

5. There exists a $\Phi \in V_0$ (called the **scaling function**) such that $\{T_\gamma \Phi\}_{\gamma \in \mathbf{Z}^n}$ form an orthonormal basis of $V_0$.

Let
$$q := |\det A| \tag{2.1}$$
so that $q$ is the index of the subgroup $A(\mathbf{Z}^n) \subset \mathbf{Z}^n$. A cross-section of the subgroup $A(\mathbf{Z}^n)$ in $\mathbf{Z}^n$ is called a set of **digits**. In other words, a set of digits is a set of coset representatives for $A(\mathbf{Z}^n)$ in $\mathbf{Z}^n$.

A set of elements $f_\gamma \in L_2(\mathbf{R}^n)$ (or more generally in any separable Hilbert space) is called a **Riesz sequence** if there exist positive constants $c$ and $C$ such that for all sequences $\{a_\gamma\}$ we have
$$c\left(\sum_\gamma |a_\gamma|^2\right)^{\frac{1}{2}} \leq \|\sum a_\gamma f_\gamma\| \leq C\left(\sum_\gamma |a_\gamma|^2\right)^{\frac{1}{2}}.$$

Suppose that $f \in L_2(\mathbf{R}^n)$ and $\hat{f}$ denotes its Fourier transform.We let
$$f_\gamma := T_\gamma f$$
so that
$$\hat{f}_\gamma(\xi) = e^{-i\langle \xi, \gamma\rangle}\hat{f}(\xi)$$
and hence
$$(\sum_\gamma a_\gamma \hat{f}_\gamma)(\xi) = a(\xi)\hat{f}(\xi), \quad a(\xi) := \sum_\gamma a_\gamma e^{-i\langle \xi, \gamma\rangle}.$$
Here $a$ is a periodic function of $\xi$ with period $2\pi$ times the dual lattice and hence by Plancherel
$$\|\sum a_\gamma f_\gamma\|^2 = \|\sum a_\gamma \hat{f}_\gamma\|^2$$
$$= \int_{\mathbf{R}^n} |a(\xi)|^2 |\hat{f}(\xi)|^2 d\xi = \int_{\mathbf{R}^n/(2\pi\mathbf{Z}^n)} |a(\xi)|^2 \sum_\gamma |\hat{f}(\xi + 2\pi\gamma)|^2 d\xi.$$
Thus we can describe when $\{f_\gamma\}$ form a Riesz sequence in terms of the periodic function $\sum_\gamma |\hat{f}(\xi + 2\pi\gamma)|^2$. Indeed, writing
$$\mathbf{T}^n := \mathbf{R}^n/(2\pi\mathbf{Z}^n)$$
so that
$$\int_{\mathbf{T}^n} |a(\xi)|^2 d\xi = (2\pi)^n \sum_\gamma |a_\gamma|^2$$
we see that:

- The $\{f_\gamma\}$ form a Riesz sequence if and only if there are positive constants $c$ and $C$ such that
$$\frac{c^2}{(2\pi)^n} \leq \sum_\gamma |\hat{f}(\xi + 2\pi\gamma)|^2 \leq \frac{C^2}{(2\pi)^n} \quad \text{a.e. .}$$

- The $\{f_\gamma\}$ form an orthonormal sequence if and only

$$\sum_\gamma |\hat{f}(\xi + 2\pi\gamma)|^2 = (2\pi)^{-n} \quad \text{a.e..}$$

- If the $\{f_\gamma\}$ form a Riesz basis of a subspace $X \subset L_2(\mathbf{R}^n)$ then there exists a function $F$ whose translates $\{F_\gamma\}$ form an orthonormal basis of $X$, indeed we define $F$ in terms of its Fourier transform $\hat{F}$ which is given by

$$\hat{F}(\xi) := \left( \sum_\gamma |\hat{f}(\xi + 2\pi\gamma)|^2 \right)^{-\frac{1}{2}} \hat{f}(\xi).$$

Suppose we have a multiresolution analysis with scaling function $\Phi$. Let us describe when a $G \in L_2(\mathbf{R}^n)$ belongs to $V_1$ in terms of its Fourier transform $\hat{G}$. The Fourier transform of $x \mapsto G(A^{-1}x)$ is given by

$$\xi \mapsto \int_{\mathbf{R}^n} G(A^{-1}x)e^{-i\langle \xi, x\rangle} dx = |\det A| \int_{\mathbf{R}^n} G(u)e^{-i\langle \xi, Au\rangle} du = q \int_{\mathbf{R}^n} G(u)e^{-i\langle A^*\xi, x\rangle} du$$

$$= q\hat{G}(A^*\xi).$$

But $G \in V_1$ if and only if $x \mapsto G(A^{-1}x)$ belongs to $V_0$ and writing this last function as $\sum a_\gamma \Phi_\gamma$ gives

$$q\hat{G}(A^*\xi) = a(\xi)\hat{\Phi}(\xi)$$

as a necessary and sufficient condition for $G$ to lie in $V_1$ where $a(\xi) := \sum_\gamma a_\gamma e^{-i\langle \xi, \gamma\rangle}$ as above. Furthermore, from

$$G(A^{-1}x) = \sum_\gamma a_\gamma \Phi_\gamma$$

and the fact that the $\Phi_\gamma$ are orthonormal it follows that

$$\sum_\gamma |a_\gamma|^2 = q \int_{\mathbf{R}^n} |G(x)|^2 dx.$$

We set

$$m_G(\xi) := q^{-1} a(\xi)$$

so that

$$\hat{G}(A^*\xi) = m_G(\xi)\hat{\Phi}(\xi).$$

Let us expand $m_G$ into a Fourier series:

$$m_G(\xi) = \sum_\gamma b(\gamma) e^{i\langle \xi, \gamma\rangle}$$

and break the sum up into a sum over $\gamma$ belonging to the various cosets of $A(\mathbf{Z}^n)$ in $\mathbf{Z}^n$ and then sum over the various cosets $E_0, \ldots, E_{q-1}$ so

$$m_G = \sum_{r=0}^{q-1} m_G^r.$$

If we choose digits (i.e. coset representatives) $\Gamma_0, \ldots, \Gamma_r$ (with $\Gamma_0 = 0$, say) we can write

$$m_G^r(\xi) = \sum_{\beta \in A(\mathbf{Z}^n)+\Gamma_r} b(\beta) e^{i\langle \xi, \beta \rangle} = e^{\langle \xi, \Gamma_r \rangle} \sum_{\gamma \in \mathbf{Z}^n} c_r(\gamma) e^{i\langle A^*\xi, \gamma \rangle}$$

where $c_r(\gamma) := b(A\gamma + \Gamma_r)$. So if we set

$$\mu_G^r := \sum_{\gamma \in \mathbf{Z}^n} c_r(\gamma) e^{i\langle A^*\xi, \gamma \rangle}$$

we have

$$m_G^r = e^{i\langle \xi, \Gamma_r \rangle} \mu_G^r.$$

We will set

$$\mu_G := \begin{pmatrix} \mu_G^0 \\ \vdots \\ \mu_G^{q-1} \end{pmatrix}$$

so that $\mu_G$ is a $\mathbf{C}^q$ valued periodic function of $\xi$.

Let $H$ be a second element of $V_1$ so we can consider its corresponding $\mathbf{C}^q$ valued function $\mu_H$.

**Proposition 2.1.1** *The collection of functions $\{G_\beta, H_\gamma\}$ forms an orthonormal system if and only if the pair of $\mathbf{C}^q$ vectors $\mu_G(\xi), \mu_H(\xi)$ is orthonormal for almost all $\xi$.*

**Proof.** We wish to show that the conditions

$$\begin{aligned} (G_{\alpha_1}, G_{\alpha_2}) &= \delta_{\alpha_1,\alpha_2} \\ (G_\alpha, H_\beta) &= 0 \\ (H_{\beta_1}, H_{\beta_2}) &= \delta_{\beta_1,\beta_2} \end{aligned}$$

(where the $(\cdot,\cdot)$ on the left hand side of these equations is the scalar product in $L_2(\mathbf{R}^n)$) are equivalent to the orthonormality of $\mu_G(\xi), \mu_H(\xi)$ for almost all $\xi$.

By Plancherel, the first of these equations becomes

$$\int_{\mathbf{R}^n} |\hat{G}(\xi)|^2 e^{i\langle \xi, \alpha_1 - \alpha_2 \rangle} d\xi = \delta_{\alpha_1,\alpha_2}.$$

Substitute $\xi = A^*\eta$ and write $\hat{G}(A^*\eta) = m_G(\eta)\hat{\Phi}(\eta)$ to obtain

$$q \int_{\mathbf{R}^n} |m_G(\eta)|^2 |\hat{\Phi}(\eta)|^2 e^{i\langle \eta, A(\alpha_1 - \alpha_2) \rangle} d\eta$$

$$= q\int_{\mathbf{T}^n} |m_G(\eta)|^2 e^{i\langle \eta, A(\alpha_1-\alpha_2)\rangle} \sum_{\gamma\in\mathbf{Z}^n} |\hat{\Phi}(\eta+\gamma)|^2 d\eta =$$

$$\frac{q}{(2\pi)^n}\int_{\mathbf{T}^n} |m_G(\eta)|^2 e^{i\langle \eta, A(\alpha_1-\alpha_2)\rangle} d\eta$$

since $\sum_{\gamma\in\mathbf{Z}^n} |\hat{\Phi}(\eta+\gamma)|^2 = \frac{1}{(2\pi)^n}$ almost everywhere. Now

$$|m_G|^2(\eta) = m_G(\eta)\overline{m_G(\eta)} = \left(\sum_r m_G^r(\eta)\right)\overline{\left(\sum_r m_G^r(\eta)\right)},$$

and so the integrand can be written as

$$\left(e^{i\langle \eta, A(\alpha_1-\alpha_2)\rangle}\sum_r m_G^r(\eta)\right)\overline{\left(\sum_r m_G^r(\eta)\right)}.$$

Now $m_G^r$ is a periodic function whose Fourier series contains only exponentials from the $r$-th coset of $A(\mathbf{Z}^n)$ in $\mathbf{Z}^n$, and multiplying by $e^{i\langle \eta, A(\alpha_1-\alpha_2)\rangle}$ which is an exponential belonging to the zero-th coset does not change this property. Hence

$$\int_{\mathbf{T}^n} \left(e^{i\langle \eta, A(\alpha_1-\alpha_2)\rangle} m_G^r(\eta)\right)\overline{m_G^s(\eta)}d\eta = 0$$

if $r \neq s$ and so our integral becomes

$$\frac{q}{(2\pi)^n}\int_{\mathbf{T}^n} e^{i\langle \eta, A(\alpha_1-\alpha_2)\rangle}\sum_{r=0}^{q-1} |m_G^r(\eta)|^2 d\eta.$$

Writing $m_G^r(\eta) = e^{i\langle \eta, \Gamma_r\rangle}\mu_G^r(A^*\eta)$ we have $|m_G^r(\eta)|^2 = |\mu_G(A^*\eta)|^2$ and restoring $\xi = A^*\eta$ the above integral becomes

$$\frac{1}{(2\pi)^n}\int_{\mathbf{T}^n} e^{i\langle \xi, \alpha_1-\alpha_2\rangle}\sum_r |\mu_G^r|^2 d\xi.$$

We have thus shown that

$$(G_{\alpha_1}, G_{\alpha_2}) = \frac{1}{(2\pi)^n}\int_{\mathbf{T}^n} e^{i\langle \xi, \alpha_1-\alpha_2\rangle}\sum_r |\mu_G^r|^2(\xi) d\xi.$$

If the $G_\alpha$ form an orthonormal set then all the Fourier coefficients of the periodic function $\sum_r |\mu_G^r|^2$ vanish except the constant term which equals one. Hence $\sum_r |\mu_G^r|^2 = 1$ a.e., and conversely.

Replacing $G$ by $H$ in the above argument shows that the $\{H_\beta\}$ form an orthonormal set if and only if $\mu_H(\xi)$ is a unit vector almost everywhere. Finally, replacing $m_G\overline{m_G}$ in the above argument by $m_G\overline{m_H}$ shows that $(G_\alpha, H_\beta) = 0$ for all $\alpha$ and $\beta$ if and only if the vectors $\mu_G(\xi)$ and $\mu_H(\xi)$ are orthogonal almost everywhere. QED

Now suppose that we have $q$ functions $G_0, \ldots, G_{q-1}$ all belonging to $V_1$ and we construct the $q \times q$ matrix valued function $\mathbf{U_G}$ of $\xi$ whose $i$-th column is $\mu_{G_i}$. We know by the above proposition that the system of functions $\{(G_i)_\gamma\}$ form an orthonormal set if and only $\mathbf{U_G}(\xi)$ is unitary for almost all $\xi$. We claim that if this matrix is unitary, then the $\{(G_i)_\gamma\}$ are actually an orthonormal *basis* of $V_1$. Indeed, if not, there will be some function $H \in V_1$ orthogonal to all the $(G_i)_\gamma$ and so all the $H_\beta$ will be orthogonal to all the $(G_i)_\gamma$ and hence by the proof of the proposition, $\mu_H(\xi)$ would be orthogonal to the $q$ orhtonormal vectors $\mu_{G_0}(\xi), \ldots, \mu_{G_{q-1}}(\xi)$ for almost all $\xi$ and hence $\mu_H(\xi) = 0$ a.e. so $\hat{H} = 0$ and hence $H = 0$ a.e.

So we have proved

**Proposition 2.1.2** *If $\mathbf{G} = \{G_0, \ldots, G_{q-1}\}$ are elements of $V_1$ then the system of functions $\{(G_i)_\gamma\}$ form an orthonormal basis of $V_1$ if and only if the matrix $\mathbf{U_G}(\xi)$ is unitary for almost all $\xi \in \mathbf{T}^n$.*

## 2.2 Wavelets

A **wavelet set** associated to the dilatation matrix $A$ is a finite collection of functions $\Psi^1, \ldots, \Psi^s$ such that the functions

$$\{U_A^j(\Psi_\gamma^s)\}$$

form an orthonormal basis of $L_2(\mathbf{R}^n)$.

Using Proposition 2.1.2 we see how to construct a wavelet set with $q-1$ elements starting with a multispectral resolution and a scaling function $G$: Let $W_0$ denote the orthogonal complement of $V_0$ in $V_1$. Then

$$W_j := (U_A)^j W_0$$

is the orthogonal complement of $V_j$ in $V_{j+1}$ of $V_j$ and hence

$$L_2(\mathbf{R}^n) = \bigoplus_{j \in \mathbf{Z}} W_j.$$

Furthermore, if $\{\Psi_{\gamma_1}^1, \ldots, \Psi_{\gamma_{q-1}}^{q-1}\}$ are an orthonormal basis of $W_0$, then their images under $U_A^j$ are an orthonormal basis of $W_j$ and hence $\Psi^1, \ldots, \Psi^s$ is a wavelet set. Now the scaling function $\Phi$ produces a vector valued function $\mu_\Phi$ which is a unit vector at all $\xi \in \mathbf{T}^n$. Call this function $\mu_0$, so in the notation of Proposition 2.1.2 we are taking $G_0 = \Phi$. Extend this to a $\mathbf{U_G}$ valued function on $\mathbf{T}^n$ by some procedure (involving some orthonormalization) and let the last $q-1$ columns of $\mathbf{U_G}$ be denoted by $\mu_1, \ldots, \mu_{q-1}$. We obtain the functions $G_1, \ldots, G_{q-1}$ which then give a wavelet set. Explicitly, the $G_i, \ \ i = 1, \ldots, q-1$ are given in terms of the $\mu_i$ via their Fourier transforms:

$$\hat{G}_i(A^*\xi) = \sum_{r=0}^{q-1} e^{i\langle \xi, \Gamma_r \rangle} \mu_i^r(A^*\xi)\hat{\Phi}(\xi).$$

The problem of construction of wavelets then breaks up into several parts:

1. The construction of the scaling function.
2. The description of smoothness and decay properties of the scaling function.
3. The choice of $\mathbf{U_G}$, hopefully in such a way that
4. The same smoothness and decay properties hold for the wavelets.

## 2.3 Characteristic scaling functions and fractals.

What are the conditions on a compact set $Q$ so that some multiple of its characteristic function $\mathbf{1}_Q$ is a scaling function for a given dilatation $A$? If $c\mathbf{1}_Q$ is to be a unit vector in $L_2(\mathbf{R}^n)$ then we must have

$$|c| = \mathrm{vol}(Q)^{-\frac{1}{2}}.$$

Let us assume that $c$ has been so chosen. We will soon see that we must have $|c| = 1$. The translate $(\mathbf{1}_Q)_\gamma$ is just the characteristic function of the set $Q + \gamma$. If $\mathbf{1}_Q$ and $(\mathbf{1}_Q)_\gamma$ are to be orthogonal for $\gamma \neq 0$ we must have

$$Q \cap (Q + \gamma) = \emptyset \quad \text{for } \gamma \neq 0$$

(where equality of sets is taken in the sense of measure theory, i.e. the two sides are equal up to a set of measure zero).

The function $\mathbf{1}_Q(A^{-1}(\cdot))$ is the characteristic function of the set $AQ$. From the fact that we want $V_{-1} \subset V_0$ and that $AQ$ is compact, we conclude that

$$\mathbf{1}_{AQ} = \sum a_\gamma \mathbf{1}_{Q+\gamma}$$

where this is a finite sum. Since the $Q + \gamma$ do not overlap, we know that all the non-zero coefficients must be equal to 1. Since the volume of $AQ$ is $q = |\det A|$ times the volume of $Q$ we know that there are exactly $q$ summands. We claim that no two of the $\gamma$'s occurring on the right can belong to the same coset of $A(\mathbf{Z}^n)$ in $\mathbf{Z}^n$. Indeed if $\gamma_1 = \gamma_2 + A\beta$ for $\gamma_1, \gamma_2$ occurring on the right hand side of this equation with non-zero coefficients, then

$$A(Q + \beta) = A(Q) + A(\beta) \supset Q + \gamma_2 + A(\beta) = Q + \gamma_1.$$

Thus $A(Q + \beta) \cap A(Q)$ has positive measure, since both $A(Q + \beta)$ and $A(Q)$ contain $Q + \gamma_1$. But this implies that $(Q + \beta) \cap Q$ has positive measure which we know is impossible. Thus the $\gamma$ occurring on the right in the expansion of $\mathbf{1}_{AQ}$ above form a set of digits. We have proved:

$$AQ = \bigcup_{i=1}^{q} (Q + \Gamma_i) \tag{2.2}$$

for some set of digits $\Gamma_1, \ldots, \Gamma_q$ where $q = |\det A|$. This is the important condition, and will allow us to show via Hutchinson's theorem that $Q$ is a fractal.

We next claim that

$$\bigcup_{\gamma \in \mathbf{Z}^n} (Q+\gamma) = \mathbf{R}^n.$$

Indeed,

$$A\left(\bigcup_{\gamma \in \mathbf{Z}^n} (Q+\gamma)\right) = \bigcup_{\gamma \in \mathbf{Z}^n} (A(Q)+A\gamma) = \bigcup_{\gamma \in \mathbf{Z}^n} (Q+\gamma)$$

since the $\Gamma_i$ in (2.2) form a set of digits. So if we let $X \subset L_2(\mathbf{R}^n)$ denote the subspace consisting of those elements whose support lies in $\bigcup_{\gamma \in \mathbf{Z}^n}(Q+\gamma)$, we have $U_A(X) = X$ and $V_0 \subset X$ from which we conclude that $V_j \subset X$ for all $j$. Since we want

$$\overline{\bigcup V_j} = L_2(\mathbf{R}^n)$$

we conclude that $X = L_2(\mathbf{R}^n)$ and hence that $\bigcup_{\gamma \in \mathbf{Z}^n}(Q+\gamma) = \mathbf{R}^n$. Conversely, if this holds, then the $\bigcup V_j$ are dense in $L_2(\mathbf{R}^n)$ since if $f \neq 0 \in L_2(\mathbf{R}^n)$ we can multiply $f$ by a scalar of absolute value one to arrange that $f$ is positive on some set of positive measure which then has an intersection with positive measure with one of the sets $A^{-j}(Q+\gamma)$ and hence is not orthogonal to $V_j$.

So we have proved that the $\{Q+\gamma\}$ must form a **tiling** of $\mathbf{R}^n$, i.e.

$$\bigcup_{\gamma \in \mathbf{Z}^n} (Q+\gamma) = \mathbf{R}^n, \quad \text{and } Q \cap (Q+\gamma) = \emptyset \quad \forall\, \gamma \neq 0.$$

This clearly implies that $\mathrm{vol}(Q) = 1$, as promised.

Now let us return to the condition (2.2). For $i = 1, \ldots q$ the maps

$$C_i : \mathbf{R}^n \to \mathbf{R}^n, \quad C_i(x) := A^{-1}(x+\Gamma_i)$$

are contractions relative to an appropriate metric on $\mathbf{R}^n$ since all the eigenvalues of $A^{-1}$ have absolute value $< 1$.

A theorem of Hutchinson says that if $C_i$ is a finite set of contractions on a complete metric space $M$ then the map

$$S \mapsto C_1(S) \cup \cdots \cup C_q(S)$$

is a contraction on the set of compact subsets of $S$ in the Hausdorff metric. In particular, this map has a unique fixed point. In our case equation (2.2) says that this fixed point is $Q$, i.e.

$$Q = \bigcup_{i=1}^{q} A^{-1}(Q+\Gamma_i). \tag{2.3}$$

As is well known, Hutchinson's theorem is a way of constructing fractal sets. Furthermore, it follows that that $Q$ is completely determined via Hutchinson's theorem by the choice of digits $\Gamma_1, \ldots, \Gamma_q$. An alternative way of describing $Q$ is as

$$Q = \{x \in \mathbf{R}^n : x = \sum_1^\infty A^{-j}\Gamma_{i_j}\},$$

which explains the term "digits". Indeed the sum on the right always converges because of the contraction property of $A^{-1}$, is compact, and satisfies (2.3).

We have shown that every set of digits determines a unique compact set $Q$ satisfying (2.2). We have not yet dealt with the issue of whether the translates of $Q$ form a tiling. Let us assume for the moment that for a given set of digits, the associated $Q + \gamma$ do form a tiling.

If we set

$$Q_i := A^{-1}(Q + \Gamma_i)$$

then we can write (2.3) as

$$Q = Q_1 \cup \cdots \cup Q_q. \tag{2.4}$$

The $Q_i$ each have volume $\frac{1}{q}$ and are disjoint, so that the functions

$$q^{\frac{1}{2}}\mathbf{1}_{Q_1}, \ldots, q^{\frac{1}{2}}\mathbf{1}_{Q_q}$$

are orthonormal, and

$$\mathbf{1}_Q = \mathbf{1}_{Q_1} + \cdots + \mathbf{1}_{Q_q}.$$

So if choose $q - 1$ linear combinations

$$\Psi^i = \sum a_i^j (q^{\frac{1}{2}}\mathbf{1}_{Q_j})$$

which are orthonormal and are orthogonal to $\mathbf{1}_{Q_1} + \cdots + \mathbf{1}_{Q_q}$, i.e. satisfy

$$a_i^1 + \cdots + a_i^q = 0$$

then we obtain a wavelet set associated to $A$. This type of wavelet set is known as a (generalized) **Haar wavelet set**. It consists of discontinuous functions of compact support.

Smoother wavelets can be obtained from the Haar wavelets by convolution, as will be seen in the next section.

The simplest example is the original example discovered by Haar - Take $n = 1$ and $A$ to be multiplication by 2, and take $\Gamma_0 = 0, \Gamma_1 = 1$ Equation(2.2) becomes

$$2Q = Q \cup Q + 1$$

The wavelet ($q - 1 = 1$ in this case) is determined by the equations

$$a_1 + a_2 = 0, \quad a_1^2 + a_2^2 = 1$$

so, up to a phase factor $a_1 = \sqrt{1/2}, a_2 = -\sqrt{1/2}$ so

$$H(t) = 1, \quad t \in [0, \frac{1}{2}), \quad H(t) = -1, \quad t \in [\frac{1}{2}, 1]$$

and $H(t) = 0$ for $t \notin [0,1]$. This is the original Haar wavelet whose scaling function is $\mathbf{1}_{[0,1]}$ and whose wavelet is function is $H$.

The information about this example can be encoded in the matrix

$$\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

in the sense that

$$\begin{pmatrix} \mathbf{1}_{[0,1]}(x) \\ H(x) \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} \mathbf{1}_{[0,1]}(2x) \\ \mathbf{1}_{[0,1]}(2x-1) \end{pmatrix}.$$

Let us now return to the logic of our situation. Given a set of digits we get a unique compact set $Q$ satisfying (2.2). We want to know for which sets of digits this set $Q$ has the property that the $Q + \gamma$ form a tiling. There are two conditions to be a tiling:

$$\bigcup (Q + \gamma) = \mathbf{R}^n$$

and

$$Q \cap (Q + \gamma) = \emptyset \quad \text{for all} \quad \gamma \neq 0.$$

We shall show that the first of these conditions always holds. So let

$$K := \bigcup_\gamma (Q + \gamma).$$

By construction, $AK = K$. Also, $K$ is closed: Indeed, suppose $x_j + \gamma_j \to z$ with $x_j \in Q$ and $\gamma_j \in \mathbf{Z}^n$. Since $Q$ is compact this implies that the $\gamma_j$ lie in a bounded, hence finite set. So by passing to a subsequence may assume that the $\gamma_j$ are constant, say all equal to $\gamma$. Hence $x_n \to z - \gamma$. But since $Q$ is compact, $z - \gamma \in Q$ so $z \in K$. So $K$ is closed. Every element of $\mathbf{R}^n$ lies within a finite distance, say $d$ of $K$, by the definition of $K$. Now let $y$ be any point of $\mathbf{R}^n$ and choose a sequence of points $x_j \in K$ each within distance $d$ of $A^j y$. Let $z_j = A^{-j} x_j$. Then

$$\|y - z_j\| = \|A^{-j}(A^j y - x_j)\| \leq C r^j d \to 0,$$

and since $K$ is closed we have $y \in K$.

Thus for any set $S$ of digits there is a unique $Q$ satisfying (2.2) with

$$\bigcup (Q + \gamma) = \mathbf{R}^n.$$

The question is - what are the conditions on $S$ so that $Q \cap (Q + \gamma) = 0$ for all $\gamma \neq 0$, or, what amounts to the same thing, that, that

$$\text{vol } Q = 1.$$

For this consider the trigonometric polynomial

$$m(\xi) := \frac{1}{q} \sum_{\Gamma \in S} e^{-i\langle \xi, \Gamma \rangle}.$$

Notice that $m(0) = 1$. Also, let $\sigma$ range over the group

$$(A^t)^{-1}(2\pi\mathbf{Z})^n/(2\pi\mathbf{Z})^n$$

(or a set of representatives of this group in $A^t(2\pi\mathbf{Z})^n$) so that

$$(\sigma, \Gamma) \mapsto e^{i\langle\sigma,\Gamma\rangle}$$

identifies $(A^t)^{-1}(2\pi\mathbf{Z})^n/(2\pi\mathbf{Z}^n)$ as the dual group to $\mathbf{Z}^n/A(\mathbf{Z})^n$. Now

$$|m(\xi)|^2 = m(\xi)\overline{m(\xi)} = \frac{1}{q^2}\sum_{\Gamma',\Gamma} e^{i\langle\xi,\Gamma'-\Gamma\rangle}$$

so

$$\sum_\sigma |m(\xi+\sigma)|^2 = \frac{1}{q^2}\sum_{\Gamma',\Gamma} e^{i\langle\xi,\Gamma'-\Gamma\rangle}\sum_\sigma e^{i\langle\sigma,\Gamma'-\Gamma\rangle}.$$

The innermost sum on the right (over $\sigma$) equals $q$ if $\Gamma' = \Gamma$ and vanishes otherwise. If we then sum over $\Gamma' = \Gamma$ we get another contribution of $q$;leading to

$$\sum_{\sigma\in(A^t)^{-1}(2\pi\mathbf{Z})^n/(2\pi\mathbf{Z})^n} |m(\xi+\sigma)|^2 = 1.$$

In section 2.5 we will show that these two properties, namely $m(0) = 1$ and $\sum_\sigma |m(\xi+\sigma)|^2 \equiv 1$ are enough to imply that the volume of the set $Q$ is an integer. This is true for any set of digits. If, in addition, $m$ satisfies a certain non-vanishing condition, see (2.19) below, $Q$ has volume one.

## 2.4 Examples of compactly supported scaling functions.

A scaling function $\Phi$, by definition, satisfies

$$\Phi(x) = \sum c_\gamma \Phi(Ax - \gamma) \tag{2.5}$$

in $L_2(\mathbf{R}^n)$, so that the sum on the right might be infinite. Suppose we look for scaling functions which are compactly supported. Then the sum on the right is finite. As $||\Phi||^2 = 1$ we must have

$$\sum |c_\gamma|^2 = q$$

and we will have found a wavelet set if we extend the row consisting of the $c$'s to a matrix with $q$ rows which are mutually orthogonal and satisfy the same condition on the sum of the squares of their absolute values.

For example, still with $n = 1$ and $A$ multiplication by 2, consider the two by four matrix

$$\begin{pmatrix} a_0(\theta) & a_1(\theta) & a_2(\theta) & a_3(\theta) \\ -a_3(\theta) & a_2(\theta) & -a_1(\theta) & a_0(\theta) \end{pmatrix}$$

where

$$\begin{aligned} a_0(\theta) &:= \tfrac{1}{2}\left(1+\sqrt{2}\cos\left[\theta+\tfrac{\pi}{4}\right]\right) \\ a_1(\theta) &:= \tfrac{1}{2}\left(1+\sqrt{2}\cos\left[\theta-\tfrac{\pi}{4}\right]\right) \\ a_2(\theta) &:= \tfrac{1}{2}\left(1-\sqrt{2}\cos\left[\theta+\tfrac{\pi}{4}\right]\right) \\ a_3(\theta) &:= \tfrac{1}{2}\left(1-\sqrt{2}\cos\left[\theta-\tfrac{\pi}{4}\right]\right) \end{aligned}$$

As a special case, if we take $\theta = \frac{\pi}{6}$ the matrix above becomes the Daubichies matrix

$$\frac{1}{4}\begin{pmatrix} 1+\sqrt{3} & 3+\sqrt{3} & 3-\sqrt{3} & 1-\sqrt{3} \\ -1+\sqrt{3} & 3-\sqrt{3} & -3-\sqrt{3} & 1+\sqrt{3} \end{pmatrix}.$$

We shall show in the next section how to associate to every such matrix a scaling function and wavelet system. But first some structural properties associated to the scaling equation (2.5): If we take the Fourier transform of the equation

$$f(x) = \sum a_\gamma f(Ax-\gamma)$$

we obtain

$$\hat{f}(\xi) = a(A^{-1*}\xi)\hat{f}(A^{-1*}\xi)$$

where

$$a(\xi) = \sum a_\gamma e^{-i\langle \xi,\gamma\rangle},$$

a finite trigonometric polynomial in our compact situation. But

$$\hat{f}(\xi) = a(A^{-1*}\xi)\hat{f}(A^{-1*}\xi) \text{ and } \hat{g}(\xi) = b(A^{-1*}\xi)\hat{g}(A^{-1*}\xi)$$
$$\Rightarrow (fg)(\xi) = c((A^{-1*}\xi))(fg)(\xi)$$

where $c = ab$ is again a trigonometric polynomial. Since the Fourier transform carries convolution into multiplication,we see that the convolution of two compactly supported solutions of an equation of type (2.5) is again a solution of an equation of type (2.5).

If $f$ and $g$ are both non-negative and of compact support, so is their convolution, and if $\int_{\mathbf{R}^n} f(x)dx = 1 = (2\pi)^{n/2}\hat{f}(0)$ and $\int_{\mathbf{R}^n} g(x)dx = 1 = (2\pi)^{n/2}\hat{g}(0)$ then $\int_{\mathbf{R}^n}(f\star g)(x)dx = 1$. But if the $f_\gamma$ form an orthonormal set as do the $g_\gamma$ it need not be true (and in general won't be true) that the $(f\star g)_\gamma$ form an orthonormal set. However, under certain circumstances, we will be able to conclude that they do form a Riesz set. So it is important for us to have the following proposition:

**Proposition 2.4.1** *Suppose that $f$ is a solution of (2.5), that $\hat{f}(0) \neq 0$ and $\hat{f}$ is continuous at $\xi = 0$. Suppose further that the $f_\gamma$ form a Riesz sequence. Let $V_0$ be the closed space spanned by the $f_\gamma$, and let*

$$V_j := U_A^j V_0.$$

*Let $P_j$ denote projection onto $V_j$. Then*

1. *For each $g \in L_2\mathbf{R}^n$ we have $\lim_{j\to-\infty} P_j g = 0$. In particular, $\bigcap V_j = \{0\}$.*
2. *$\bigcup V_j$ is dense in $L_2\mathbf{R}^n$.*
3. *The $V_j$ form a multiresolution analysis.*

**Proof.** We know that we can replace the function $f$ by a function $F$ lying in $V_0$ with the $F_\gamma$ orthonormal. In other words we can replace $f$ an $F$ which is a scaling function for the same family $V_j$. So 3.) follows from 1.) and 2.).

To prove 1.) it is enough to take $g$ of compact support, since the functions of compact support are dense in $L_2\mathbf{R}^n$. Let $K$ denote the support of $g$, and let

$$f_{j\gamma}(x) := |\det A|^{\frac{j}{2}} f(A^j x - \gamma).$$

Let $J$ be such that the sets $A^j K - \gamma$ are disjoint for all $j \leq J$. For each $j$ the $f_{j\gamma}$ form a Riesz basis of $V_j$ and so

$$\|P_j g\|^2 \leq c^{-1} \sum_\gamma |(P_j g, f_{j\gamma})|^2 = c^{-1} \sum_\gamma |(g, f_{j\gamma})|^2 = c^{-1} \sum_\gamma \int_K g(x) \overline{f_{j\gamma}(x)} dx.$$

By the Cauchy Schwarz inequality we this last sum is $\leq$

$$B \sum_\gamma \int_K |f_{j\gamma}(x)|^2 dx, \quad \text{where } B = c^{-1}\|g\|^2.$$

Now

$$\int_K |f_{j\gamma}(x)|^2 dx = |\det A|^j \int_K |f(A^j x - \gamma)|^2 dx = \int_{A^j K - \gamma} |f(u)|^2 du.$$

For $j \leq J$ the sets $A^j K - \gamma$ are disjoint, so

$$\|P_j g\|^2 \leq \int_{Y_j} |f(x)|^2 dx, \quad \text{where } Y_j := \bigcup (A^j K - \gamma).$$

By the Lebesgue dominated convergence theorem this tends to zero. This completes the proof of 1. Notice that in this proof we only made use of hypothesis that the $f_\gamma$ form a Riesz sequence.

Next, suppose that $h$ is orthogonal to $\bigcup V_j$, i.e. that $P_j h = 0$ for all $j$. We must show that $h = 0$ to prove ii). Find a compact set $K$ such that $\int_{K^c} |\hat{h}(\xi)|^2 dx < \epsilon^2$ which is possible since $h \in L_2$. Define $g$ by $\hat{g} = \mathbf{1}_K \cdot \hat{h}$ (where now $K$ is a compact subset of $\xi$ space) so that we have $\|h - g\| \leq \epsilon$. We will show that for any such choice we have

$$\|g\|^2 \leq \frac{\epsilon^2}{C(2\pi)^n |\hat{f}(0)|}. \tag{2.6}$$

If we choose $\epsilon$ so that $\epsilon/C[(2\pi)^n|\hat{f}(0)|] < 1$ we conclude that $\|g\| < \epsilon$ and hence $\|h\| < 2\epsilon$. Since $\epsilon$ can be chosen (positive but) arbitrarily small, this will prove that $h = 0$. So we must prove (2.6). We have

$$C\|P_j g\|^2 \geq \sum_\gamma |(g, f_{j\gamma})|^2 = \sum_\gamma \left| \int \hat{g}(\xi)\overline{(f_{j\gamma})\hat{}(\xi)} d\xi \right|^2.$$

Now

$$(f_{j\gamma})\hat{}(\xi) = |\det A|^{-j/2} e^{-i\langle \xi, A^{-j}\gamma\rangle} \hat{f}(A^{-j}(\xi)).$$

Suppose we choose $j$ so large that $K \subset A^j[-\pi,\pi]^n$. Then each of the summands is just the Fourier coefficient of $\hat{g}(\xi)\hat{f}(A^j\xi)$ (up to some overall constants. We conclude that

$$C\|P_j g\|^2 \geq (2\pi)^n \int_K |\hat{g}\overline{\hat{f}(A^{-j}\xi)}|^2 d\xi.$$

But $\hat{f}(A^{-j}\xi) \to \hat{f}(0)$ uniformly on $K$ as $j \to \infty$. QED

## 2.5 Product Fourier expansion.

Let us return to equation (2.5) which we will write in Fourier transform language as

$$\hat{\Phi}(\xi) = m((A^t)^{-1}\xi)\hat{\Phi}((A^t)^{-1}\xi) \tag{2.7}$$

where

$$m(\xi) := |\det A|^{-1} \sum_\gamma c_\gamma e^{-i\langle \xi, \gamma \rangle}. \tag{2.8}$$

If we set $\xi = 0$ in (2.7) (and assume $\hat{\Phi}(0) \neq 0$) then it follows that we must have

$$m(0) = 1. \tag{2.9}$$

It is then clear that (up to a scalar factor) $\hat{\Phi}$ must be equal to the infinite product

$$\prod_{j=1}^{\infty} m((A^t)^{-j}\xi). \tag{2.10}$$

Since $m$ is a finite trigonometric polynomial, it is differentiable at $\xi = 0$, and since $m(0) = 1$ we have

$$|m(\xi) - 1| < C|\xi|$$

for some constant $C$ and all $\xi$ (since $m$ is periodic). Hence

$$|m((A^t)^{-j}\xi) - 1| < Cr^j|\xi|$$

for some $0 < r < 1$ since $(A^t)^{-1}$ is a linear contraction. This shows that the infinite product converges for all $\xi$, uniformly on compact sets. We claim that $\hat{\Phi}$ has polynomial growth in $\xi$. These two facts then show that $\Phi$ exists as a distribution. To see the polynomial growth, let $\xi$ be any point such that $\|\xi\| \geq 1$ and choose $\ell = \ell(\xi)$ such that

$$r^{\ell}\|\xi\| \leq 1 \leq \|\xi\|,$$

in particular

$$\ell \leq 1 - \frac{\log \|\xi\|}{\log r}.$$

Break the infinite product for $\hat{\Phi}$ into two parts:

$$\hat{\Phi}(\xi) = \prod_{1}^{\ell} m((A^t)^{-1}\xi) \prod_{\ell}^{\infty} m((A^t)^{-1}\xi).$$

The second product converges uniformly in $\xi$ and hence can be estimated in absolute value by some constant, $R$ independently of $\xi$. Since the function $m$ is periodic and continuous,it is bounded, so the first (finite) product above can be estimated in absolute value by $k^m$ where $k \geq \sup_\eta |m(\eta)|$ and we choose $k > 1$. Thus

$$|\hat{\Phi}(\xi)| \leq Rk^{\ell} \leq Ak\left(k^{-\frac{\log \|\xi\|}{\log r}}\right).$$

But

$$k^{-\frac{\log \|\xi\|}{\log r}} = \|\xi\|^{-\frac{\log k}{\log r}}$$

as can be seen by taking logarithms of both sides. Thus

$$|\hat{\Phi}(\xi)| \leq Rk\|\xi\|^{N}, \quad N = -\frac{\log k}{\log r}.$$

Furthermore similar estimates hold uniformly on the finite products converging to $\hat{\Phi}$. This shows that $\Phi$ exists as a distribution.

We will now impose additional conditions on $m$ which will imply that $\hat{\Phi}$ and hence $\Phi$ belong to $L_2$.

For this, let $D$ be a fundamental domain of the lattice $\mathbf{Z}^n$ in $\mathbf{R}^n$, let

$$\Phi_0 := \mathbf{1}_D$$

and define

$$\Phi_{j+1}(x) := \sum_{\gamma} c_\gamma \Phi_j(Ax - \gamma)$$

so

$$\hat{\Phi}_j(\xi) := \hat{\Phi}_0((A^t)^{-j}\xi) \prod_{1}^{j} m((A^t)^{-k}\xi).$$

Thus $\Phi_j$ is a sequence of functions which converge to $\Phi$ in the distributional topology. If we can show that

$$\|\Phi_j\| = 1$$

for all $j$, then we can choose a subsequence of the $\Phi_j$ which converge in the weak topology to some element of $L_2$, since the unit ball in a separable Hilbert space is compact in the weak topology. But then this limit must equal $\Phi$.

For $u, v \in L_2(\mathbf{R}^n)$ define their lattice correlation $\mathbf{corr}(u,v)$, a function on the lattice $\mathbf{Z}^n$ by

$$\mathbf{corr}(u,v)(\gamma) = (u, v(\cdot - \gamma)) = \int_{\mathbf{R}^n} u(x)\overline{v(x-\gamma)}dx.$$

We can also write this as

$$\mathbf{corr}(u,v)(\gamma) = u \star \tilde{v}(\gamma). \tag{2.11}$$

Let us write the scaling equation (2.5) as

$$\Phi = C\Phi$$

where $C$ is the operator

$$(Cu)(x) = (c \star u)(Ax)$$

where we now think of $c$ as a distribution - as a sum of delta functions:

$$c = \sum_\gamma c_\gamma \delta_\gamma.$$

We now compute $\mathbf{corr}(Cu, Cv)$ in terms of $\mathbf{corr}(u,v)$ and will see that it is given by "matrix multiplication". For this it will be convenient to introduce some notation: Set

$$\mathbf{p} := c \star \tilde{c} \tag{2.12}$$

so, in a sense, $\mathbf{p} = \mathbf{corr}(c,c)$, it is the "auto-correlation" of $c$ and we will also write

$$\mathbf{p} = \mathbf{aut}(c)$$

to emphasize this point. Let us also define the matrix

$$\mathbf{T}$$

indexed by elements of $\mathbf{Z}^n$ by

$$\mathbf{T}_{\alpha\beta} := |\det A|^{-1}\mathbf{p}(A\alpha - \beta). \tag{2.13}$$

We have

$$\begin{aligned}
\mathbf{corr}(Cu, Cv)(\alpha) &= ((c \star u)(A\cdot), (c \star u)(A \cdot - A\alpha)) \\
&= |\det A|^{-1}\,((c \star u)(\cdot), (c \star u)(\cdot - A\alpha)) \\
&= |\det A|^{-1}(c \star u) \star (c \star v)\tilde{}\,(A\alpha) \\
&= |\det A|^{-1}(c \star \tilde{c}) \star (u \star \tilde{v})(A\alpha) \\
&= |\det A|^{-1}\mathbf{p} \star \mathbf{corr}(u,v)(A\alpha) \\
&= \sum_\beta \mathbf{T}_{\alpha\beta}\,\mathbf{corr}(u,v)(\beta).
\end{aligned}$$

We have proved

$$\mathbf{corr}(Cu, Cv) = \mathbf{T}\,\mathbf{corr}(u, v). \tag{2.14}$$

Now the assertion that the $\Phi_\gamma$ be orthonormal is the same as the assertion that

$$\mathbf{corr}(\Phi, \Phi) = \delta,$$

the delta function at the origin of $\mathbf{Z}^n$. So if in addition $C\Phi = \Phi$, we must have

$$\mathbf{T}\delta = \delta.$$

This says that the delta function on $\mathbf{Z}^n$ is an eigenvector of $\mathbf{T}$ with eigenvalue 1, or, what amounts to the same thing, that the "central column" of $\mathbf{T}$ is $\delta$, i.e. we must require

$$\mathbf{aut}(c)(A\gamma) = |\det A|\delta(\gamma). \tag{2.15}$$

Then (2.14) implies inductively that

$$\mathbf{corr}(\Phi_i, \Phi_i) = \delta,$$

in particular that $\|\Phi_i\| = 1$ and hence that $\Phi \in L_2(\mathbf{R}^n)$.

We can express condition (2.15) in terms of $m$: The Fourier series corresponding to $\mathbf{aut}c$ is $|\det A|^2|m(\xi)|^2$ since the Fourier series $\sum_\gamma \overline{c}_{-\gamma}e^{-i\langle\xi,\gamma\rangle}$ is just $\overline{m(\xi)}$ and the Fourier series takes convolution into multiplication. Now the dual group to the sublattice is the quotient group of the torus $\mathbf{R}^n/(2\pi\mathbf{Z})^n$ by the subgroup

$$(A^t)^{-1}(2\pi\mathbf{Z})^n$$

and the surjection dual to the injection of $A\mathbf{Z}^n \to \mathbf{Z}^n$ is averaging over the action of the quotient

$$G_A := (A^t)^{-1}(2\pi\mathbf{Z})^n)/(2\pi\mathbf{Z})^n.$$

Since the Fourier series of the delta function at the origin (for the sublattice) is the constant one (one the dual group) we see that (2.15) is equivalent to

$$\sum_{\sigma\in G_A} |m(\sigma+\xi)|^2 = 1 \tag{2.16}$$

We can verify the equivalence of (2.16) and (2.15) directly without an appeal to group theory. Let

$$\hat{c}(\xi) := \sum c_\gamma e^{-i\langle\xi,\gamma\rangle}$$

so that

$$|m(\xi)|^2 = |\det A|^{-2}|c(\xi)|^2 = q^{-2}|c(\xi)|^2$$

and

$$|c(\xi)|^2 = \sum_\gamma \mathbf{aut}(c)(\gamma)e^{-i\langle\xi,\gamma\rangle}.$$

Then

$$\begin{aligned}\sum_{\sigma\in G_A}|c(\xi+\sigma)|^2 &= \sum_{\sigma\in G_A}\sum_{\gamma}\mathbf{aut}(c)(\gamma)e^{-i\langle\xi+\sigma,\gamma\rangle}\\ &= \sum_{\gamma}\mathbf{aut}(c)(\gamma)e^{-i\langle\xi,\gamma\rangle}\sum_{\sigma\in G_A}e^{-i\langle\sigma,\gamma\rangle}\\ &= q\sum_{\gamma}\mathbf{aut}(c)(\gamma)\mathbf{1}_{A\mathbf{Z}^n}(\gamma)e^{-i\langle\xi,\gamma\rangle}\end{aligned}$$

where $\mathbf{1}_{A\mathbf{Z}^n}$ denotes the characteristic function of the sublattice $A\mathbf{Z}^n$, i.e. is one on this sublattice and zero elsewhere on $\mathbf{Z}^n$. The last sum above is just the Fourier series of the sequence $\mathbf{aut}(c)\mathbf{1}_{A\mathbf{Z}^n}$. So the last expression above is identically equal to $q$ if and only if (2.15) holds. So (2.15) is equivalent to (2.16).

Notice that $m(0) = 1$ and (2.16) implies that $m(\sigma) = 0$ for $\sigma \neq 0$ in $G_A$. In other words, $m(\xi) = 0$ for $\xi \in (A^t)^{-1}(2\pi\mathbf{Z})^n$ but not in $(2\pi\mathbf{Z})^n$. Suppose that $\xi \neq 0$, $\xi \in (2\pi\mathbf{Z})^n$. Then some term on the right of the infinite product expansion for $\hat{\Phi}$ must vanish. Thus

$$\hat{\Phi}(\tau) = 0, \quad \tau \in (2\pi\mathbf{Z})^n, \quad \tau \neq 0.$$

But the Poisson summation formula says that

$$\sum_{\gamma} f(x-\gamma) = (2\pi)^{n/2}\sum_{\tau\in(2\pi\mathbf{Z})^n}\hat{f}(\tau)e^{-i\langle\tau,x\rangle}.$$

Taking $f = \Phi$ all the coefficients on the right vanish except when $\tau = 0$ where it equals $\hat{\Phi}(0)$ so

$$\sum_{\gamma}\Phi(x-\gamma) \equiv (2\pi)^{n/2}\hat{\Phi}(0). \tag{2.17}$$

For example, we have already seen that if we take $c(\gamma)$ to be one when $\gamma$ belongs to a set of digits and zero elsewhere, then $c$ satisfies (2.15) as we have verified that that $m$ satisfies (2.16). If $Q$ is the set determined by (2.2), we have verified that $\Phi = \mathbf{1}_Q$ satisfies (2.5). But since $\mathbf{1}_Q$ takes on the values 0 or 1 only, we see that the left hand side of (2.17) is an integer, and this integer is independent of $x$ by (2.17). On the other hand $\hat{\Phi}(0) = (2\pi)^{-n/2}\text{vol }(Q)$ so the right hand side of (2.17) is vol $(Q)$. We have shown that in all cases the volume of the set $Q$ given by (2.2) is an integer, as promised. Shortly we will return to the question of when this integer equals one.

In the meanwhile let us return to general considerations. We claim that

$$\text{supp } \Phi \text{ is compact.} \tag{2.18}$$

To see this, observe that $m$ is the Fourier transform of a signed measure concentrated on a compact (in fact finite) set, $S$. Hence $m((A^t)^{-j}\cdot)$ is the Fourier transform of a signed measure concentrated on the set $A^{-j}$ and hence the support of $\Phi$ is contained in the set

$$S + A^{-1}S + A^{-2}S + \cdots$$

which is bounded. QED.

We now turn to Cohen's theorem which gives a (necessary and) sufficient condition on a set of digits $S$ to guarantee that $|Q| = 1$.

For the trigonometric polynomial $m$ that enters into (2.16), let $\nu_N$ denote the finite product

$$\nu_N(\xi) := \prod_1^N m((A^t)^{-j}\xi)$$

so that

$$\hat{\Phi}(\xi) = \nu_N(\xi)\hat{\Phi}((A^t)^{-N}\xi)$$

for any solution of (2.5), and, up to a constant the limit of the $\nu_N$ as $N \to \infty$ is $\Phi$. Since $m(0) = 1$ and, for $\Phi = \mathbf{1}_Q$ we have verified that $\hat{\Phi}(0) = (2\pi)^{-n/2}|Q|$ we conclude that

$$\nu_N(\xi) \to \frac{(2\pi)^{n/2}}{|Q|}\hat{\Phi}(\xi)$$

the convergence being pointwise as $N \to \infty$.

On the other hand, let $D$ be a fundamental domain for $\mathbf{Z}^n$, let $\Phi_0 = \mathbf{1}_D$ and $\Phi_{j+1}(\xi) = \sum_{\gamma\in S} c_\gamma \Phi_j(A\xi - \gamma)$ as above, in the proof that $\Phi \in L_2$. We have proved that $\|\Phi_N\| = 1$. On the other hand

$$\hat{\Phi}_N(\xi) = \nu_N(\xi)\hat{\Phi}_0((A^t)^{-N}\xi).$$

So

$$1 = \int_{\mathbf{R}^n} |\nu_N(\xi)|^2|\hat{\Phi}_0((A^t)^{-j}\xi)|^2.$$

Since $\hat{\Phi}_0(0) = (2\pi)^{-n/2}|D| = (2\pi)^{-n/2}$, the integrand on the right hand side of this equation tends pointwise to

$$\frac{1}{|Q|^2}|\hat{\Phi}|^2.$$

If passage to the limit under the integral sign were legitimate, we would conclude form this (and Plancherel's theorem) that

$$\|\Phi\|^2 = |Q|^2.$$

But by definition, $\|\Phi\|^2 = \int_{\mathbf{R}^n} |\mathbf{1}_Q|^2 dx = |Q|$. So $|Q|^2 = |Q|$ and since $|Q| \neq 0$ this would imply that $|Q| = 1$. So we must make some hypothesis which will legitimize passing to the limit under the integral:

Let $\Omega$ be a fundamental domain for $(2\pi\mathbf{Z})^n$, for example, $\Omega = [-\pi, \pi]^n$. Then

$$\int_{\mathbf{R}^n} |\nu_N(\xi)|^2|\hat{\Phi}_0((A^t)^{-j}\xi)|^2 = \sum_{\gamma\in\mathbf{Z}^n}\int_{B^N\Omega+2\pi B^N\gamma} |\nu_N(\xi)|^2|\hat{\Phi}_0(B^{-N}\xi)|^2 d\xi$$

where we have written $B$ for $A^t$. Since $\nu_N$ is $B^N(2\pi\mathbf{Z})^n$ periodic, we can write this last sum as

$$\int_{B^N\Omega} |\nu_N(\xi)|^2 \sum_\gamma |\hat{\Phi}_0(B^{-N}\xi - 2\pi\gamma)|^2 d\xi.$$

But since the $\Phi_0(\cdot - \gamma)$ form an orthonormal sequence, we know that

$$\sum_\gamma |\hat{\Phi}_0(\eta - 2\pi\gamma)|^2 \equiv \frac{1}{(2\pi)^n}.$$

Hence

$$1 = \frac{1}{(2\pi)^n} \int_{B^N\Omega} |\nu_N(\xi)|^2 d\xi.$$

This formula has been derived for a choice $\Omega$ of fundamental domain for $(2\pi\mathbf{Z})^n$. But since $\nu_N$ is $(2\pi N\mathbf{Z})^n$ periodic, we can replace $\Omega$ by any other fundamental domain, $K$. So we have

$$\int_{\mathbf{R}^n} |\nu_N(\xi)|^2 \mathbf{1}_{B^N K}(\xi) d\xi = (2\pi)^n.$$

Suppose that $K$ contains a neighborhood of the origin. Then $\mathbf{1}_{B^N K}(\xi) \to 1$ for all $\xi$, and the integrand above converges pointwise to

$$\frac{(2\pi)^n}{|Q|^2} |\hat{\Phi}(\xi)|^2$$

and if we could justify passing to the limit under the integral sign we would conclude that $|Q| = 1$ by the argument given above.

Suppose that $K$ satisfies

$$m((A^t)^{-j}\xi) \neq 0 \quad \forall \xi \in K. \tag{2.19}$$

Then $\nu_N(\xi) \neq 0 \ \ \forall \xi \ \in K$ and since $\hat{\Phi}(0) \neq 0$ and

$$\hat{\Phi}(\xi) = \nu_N(\xi)\Phi(B^{-N}\xi)$$

we conclude that

$$\hat{\Phi}(\xi) \neq 0, \quad \forall\ \xi \in K.$$

So there is some constant $c$ such that $|\Phi(\xi)| > c$ for $\xi \in K$ and hence that $|\Phi(B^{-N}\xi)| > c$ for $\xi \in B^N K$ and hence from

$$\hat{\Phi}(\xi) = \nu_N(\xi)\Phi(B^{-N}\xi)$$

once again that

$$|\nu_N(\xi)| \leq c^{-1}|\hat{\Phi}(\xi)| \quad \forall \xi \in B^N K$$

or

$$|\nu_N(\xi)|^2 \mathbf{1}_{B^N K}(\xi) \leq C^{-2}|\hat{\Phi}(\xi)|^2$$

for all $\xi \in \mathbf{R}^n$ But $\hat{\Phi} \in L_2(\mathbf{R}^n)$ so we can apply the Lebesgue dominated convergence theorem to justify taking the limit under the integral sign. We have proved the follow theorem of Cohen:

**Proposition 2.5.1** *If there exists a fundamental domain $K$ for $(2\pi\mathbf{Z})^n$ which contains a neighborhood of the origin and satisfies (2.19) then $|Q| = 1$.*

# Chapter 3

# Wavelets on R.

## 3.1 Desiderata.

We redo some of the results of the preceding chapter for the special one dimensional case in order to tie in with notations is various books. Let us go back to the scaling equation for $\mathbf{R}$ which we write as

$$\phi(t) = \sqrt{2}\sum_k c(k)\phi(2t-k) \tag{3.1}$$

and let $\mathbf{c}$ denote the vector with coefficients $c(k)$. We set

$$\mathbf{h} = \frac{\mathbf{1}}{\sqrt{\mathbf{2}}}\mathbf{c}$$

and also think of $\mathbf{h}$ as the distribution

$$\mathbf{h} = \sum h(k)\delta_k$$

so that the scaling equation becomes

$$\phi(t) = 2[\mathbf{h} \star \phi](2t) \tag{3.2}$$

where we are using the notation

$$(f \star g)(x) = \int_{\mathbf{R}} f(y)g(x-y)dy$$

for convolution. Integrating both sides of

$$\phi(t) = 2\sum_k h(k)\phi(2t-k) \tag{3.3}$$

gives

$$\sum h(k) = 1$$

(if we want $\int_{\mathbf{R}} \phi(t)dt \neq 0$). Let

$$H(\xi) = \sum_k h(k)e^{-ik\xi}$$

so

$$H(0) = 1. \tag{3.4}$$

Taking the Fourier transform of (3.2) gives

$$\hat{\phi}(\xi) = H\left(\frac{\xi}{2}\right)\hat{\phi}\left(\frac{\xi}{2}\right). \tag{3.5}$$

The condition that the $\phi(\cdot - k)$ form an orthonormal basis of $L_2(\mathbf{R})$ translates by Fourier transform into the condition

$$\sum_\ell |\hat{\phi}(\xi + 2\pi\ell)|^2 \equiv \frac{1}{2\pi}. \tag{3.6}$$

Breaking the sum on the right into even and odd $\ell$ and then using (3.5) with $\xi$ replaced by $\frac{\xi}{2}$ and $\frac{\xi}{2} + \pi$ gives

$$\begin{aligned}
\frac{1}{2\pi} &\equiv \sum_k |\hat{\phi}(\xi + 4\pi k)|^2 + \sum_k |\hat{\phi}(\xi + 2\pi + 4\pi k)|^2 \\
&= \sum_k \left|H\left(\frac{\xi}{2}\right)\right|^2 \left|\hat{\phi}\left(\frac{\xi}{2} + 2\pi k\right)\right|^2 + \sum_k \left|h\left(\frac{\xi}{2} + \pi\right)\right|^2 \left|\hat{\phi}\left(\frac{\xi}{2} + \pi + 2\pi k\right)\right|^2 \\
&= \left(\left|H\left(\frac{\xi}{2}\right)\right|^2 + \left|H\left(\frac{\xi}{2} + \pi\right)\right|^2\right) \cdot \frac{1}{2\pi}.
\end{aligned}$$

We conclude that we must have

$$|H(\xi)|^2 + H(\xi + \pi)|^2 \equiv 1 \tag{3.7}$$

Setting $\xi = 0$ and using (4.2) gives

$$H(\pi) = 0. \tag{3.8}$$

Suppose that $H$ is a trigonometric polynomial. We know that the scaling equation in Fourier form then gives rise to an infinite product expansion for the Fourier transform of $\phi$:

$$\hat{\phi}(\xi) = \prod H\left(\frac{\xi}{2^k}\right) \tag{3.9}$$

and that this product converges as a distribution. We will want to impose further conditions which will guarantee the existence of this limit as an element of $L_2$.

Suppose we had any element $f \in V_1$. Writing

$$f = \sum f(k)\phi_{1,k}$$

and taking the Fourier transform, we obtain

$$\hat{f}(\xi) = F\left(\frac{\xi}{2}\right)\hat{\phi}\left(\frac{\xi}{2}\right) \tag{3.10}$$

where

$$F(\xi) := \frac{1}{\sqrt{2}}\sum f(k)e^{-ik\xi}.$$

Conversely, any function satisfying this last equation belongs to $V_1$. If $f \in W_0$ which is the orthogonal complement of $V_0$ in $V_1$ then

$$0 = \int_{\mathbf{R}} \hat{f}(\xi)\overline{\hat{\phi}(\xi)}e^{ik\xi}d\xi = \int_0^{2\pi}\left(\sum_k \hat{f}(\xi+2\pi k)\overline{\hat{\phi}(\xi+2\pi k)}\right)e^{ik\xi}d\xi$$

implying that the periodic function

$$\sum_k \hat{f}(\xi+2\pi k)\overline{\hat{\phi}(\xi+2\pi k)}$$

vanishes (almost everywhere). Substituting (3.10) and summing over even and odd and using (3.6) we conclude that the vector

$$\begin{pmatrix} F(\xi) \\ F(\xi+\pi) \end{pmatrix}$$

must be orthogonal to

$$\begin{pmatrix} H(\xi) \\ H(\xi+\pi) \end{pmatrix}$$

and hence there exists a function $\xi \mapsto \lambda(\xi)$ such that

$$\begin{pmatrix} F(\xi) \\ F(\xi+\pi) \end{pmatrix} = \lambda(\xi)\begin{pmatrix} \overline{H(\xi+\pi)} \\ -\overline{H(\xi)} \end{pmatrix} \tag{3.11}$$

where, in fact

$$\lambda(\xi) = F(\xi)H(\xi+\pi) - F(\xi+\pi)H(\xi)$$

obtained by taking the scalar product. The function $\lambda$ satisfies

$$\lambda(\xi+\pi) = -\lambda(\xi)$$

and hence there is a function $\nu$ periodic of period $2\pi$ such that

$$\lambda(\xi) = e^{i\xi}\nu(2\xi).$$

Substituting this into (3.11) we obtain

$$F(\xi) = e^{i\xi}\nu(2\xi)\overline{H(\xi+\pi)}$$

from the first row. Substituting this back into (3.10) gives

$$\hat{f}(\xi) = e^{i\xi/2}\overline{H\left(\frac{\xi}{2}+\pi\right)}\hat{\phi}\left(\frac{\xi}{2}\right)$$

as the condition that $f \in W_0$. In particular, we may choose $\nu \equiv 1$ and set

$$\hat{\psi}(\xi) := e^{i\xi/2}\overline{H\left(\frac{\xi}{2}+\pi\right)}\hat{\phi}\left(\frac{\xi}{2}\right). \tag{3.12}$$

Then a direct check (summing over even and odd) shows that

$$\sum_k |\hat{\psi}(\xi+2\pi k)|^2 = \frac{1}{2\pi}$$

so that the $\psi(\cdot - k)$ form an orthonormal system in $W_0$. Since every element of $f \in W_0$ has property that $\hat{f} = \nu\hat{\psi}$ for some periodic function $\nu$, expanding $\nu$ into a Fourier series shows that

$$f = \sum \nu(k)\psi(\cdot - k)$$

so the $\psi(\cdot - k)$ form an orthonormal basis of $W_0$.

Expanding out

$$\begin{aligned} e^{i\xi/2}\overline{H\left(\frac{xi}{2}+\pi\right)} &= \sum \overline{h_k} e^{ik(\frac{\xi}{2}+\pi} e^{\xi/2} \\ &= \sum (-1)^k \overline{h_k} e^{i(k+1)\xi/2} \\ &= \sum_\ell (-1)^{\ell-1} \overline{h_{-\ell-1}} e^{-i\ell\xi/2} \end{aligned}$$

we obtain

$$\hat{\psi}(\xi) = \sum (-1)^k \overline{h_{-k-1}} e^{-ik\xi/2}\hat{\phi}\left(\frac{\xi}{2}\right)$$

or

$$\psi(t) = \sum (-1)^{k-1}\overline{h_{-k-1}}\phi(2t-k).$$

We will be interested in constructing suitable $\mathbf{h}$ supported on the integers from 0 to $N$. where $N$ is some odd integer.

## 3.2 Matrix formulation

Assume that we have our vector $\mathbf{h}$ with $h(k) = 0$ unless $0 \le k \le N$ and that conditions (4.2) and (4.3) hold which say

$$\sum h(k) = 1$$

and

$$\sum(-1)^k h(k) = 0$$

or

$$h(0) + h(2) + \cdots = h(1) + h(3) + \cdots$$

and both sides equal $\frac{1}{2}$.

Define the bi-infinite matrix $\mathbf{M} = \mathbf{M}(\mathbf{h})$ by

$$\mathbf{M}_{ij} = 2h(2i - j). \tag{3.13}$$

We also let $\mathbf{m}$ denote the $N \times N$ "central block" of $\mathbf{M}$, so

$$\mathbf{m} = \begin{pmatrix} 2h(0) & 0 & 0 & \cdots \\ 2h(2) & 2h(1) & 2h(0) & \cdots \\ 2h(4) & 2h(3) & 2h(2) & \cdots \\ \vdots & \vdots & \vdots\cdots & \end{pmatrix}.$$

For example, for $N = 7$ this matrix is

$$\mathbf{m} = \begin{pmatrix} 2h(0) & 0 & 0 & 0 & 0 & 0 & 0 \\ 2h(2) & 2h(1) & 2h(0) & 0 & 0 & 0 & 0 \\ 2h(4) & 2h(3) & 2h(2) & 2h(1) & 2h(0) & 0 & 0 \\ 2h(6) & 2h(5) & 2h(4) & 2h(3) & 2h(2) & 2h(1) & 2h(0) \\ 0 & 2h(7) & 2h(6) & 2h(5) & 2h(4) & 2h(3) & 2h(2) \\ 0 & 0 & 0 & 2h(7) & 2h(6) & 2h(5) & 2h(4) \\ 0 & 0 & 0 & 0 & 0 & 2h(7) & 2h(6) \end{pmatrix}.$$

The columns of the matrix $\mathbf{M}$ and of the matrix $\mathbf{m}$ add up to one, so the infinite row vector of all ones is a left eigenvector of $M$ and the finite row vector of all ones is a left eigenvector for $\mathbf{m}$ with eigenvalue one.

For any function $v$ recall hat swe have defined $\tilde{v}$ by

$$\tilde{v}(t) = \overline{v(-t)}$$

with the corresponding definition for distributions so that

$$\hat{\tilde{v}} = \overline{\hat{v}}$$

and

$$[u \star \tilde{v}](y) = (u, R_y v) = \int_{\mathbf{R}} u(x)\overline{v(x-y)}dy.$$

Also,

$$\tilde{f} \star \tilde{g} = (f \star g)\tilde{}.$$

Define the operator $\mathbf{C}$ by

$$[Cu](t) := 2[\mathbf{h} \star u](2t) \tag{3.14}$$

so that the scaling equation (3.2) is

$$\phi = \mathbf{C}\phi.$$

For $u, v \in L_2(\mathbf{R})$ recall that $\mathbf{corr}(u, v)$ is the vector whose components are the correlations at the integers $k$:

$$\mathbf{corr}(u, v)(k) := (u(\cdot), v(\cdot - k)) = [u \star \tilde{v}](k).$$

Then recall that

$$\mathbf{corr}(\mathbf{C}u, \mathbf{C}v)(k) = (2[\mathbf{h} \star u](2\cdot), 2[\mathbf{h} \star v](2\cdot - 2k)) = 2([\mathbf{h} \star u](\cdot), [\mathbf{h} \star v](\cdot - 2k)).$$

We have written this last expression as

$$2[\mathbf{h} \star \tilde{\mathbf{h}}] \star (u \star \tilde{v})(2k)$$

and ven more succinctly as

$$\mathbf{corr}(\mathbf{C}u, \mathbf{C}v) = \mathbf{T}\,\mathbf{corr}(u, v) \tag{3.15}$$

where $\mathbf{T}$ is the matrix given by

$$\mathbf{T}_{ij} = 2p(2i - j)$$

where

$$\mathbf{p} = \mathbf{h} \star \tilde{\mathbf{h}}.$$

Taking the Fourier transform of this last equation gives

$$P(\xi) = |H(\xi)|^2$$

so

$$P(0) = 1, \quad \text{and} \quad P(\pi) = 0$$

implies that the columns of $\mathbf{T}$ add up to one. Also, since $\mathbf{h} \star \tilde{\mathbf{h}}$ has support on the interval $[-N, N]$ and so the columns of $\mathbf{T}$ are finitely supported and we will want to consider the "central block" of size $(2N - 1) \times (2N - 1)$ of $\mathbf{T}$ which we shall denote by $\mathbf{t}$. It also has the row vector of all ones as an eigenvector with eigenvalue one.

If $\phi$ is a solution of the scaling equation, so $\mathbf{C}\phi = \phi$, it follows from (3.15) that its autocorrelation $\mathbf{corr}(\phi, \phi)$ is an eigenvector of $\mathbf{T}$ with eigenvalue 1. If $u$ is a function such that $\mathbf{corr}(u, u)$ has support in $[-N, N]$, then it follows from (3.15) that $Cu$ has this same property, and that the non-vanishing portion of $\mathbf{corr}(Cu, Cu)$ is obtained from that of $\mathbf{corr}(u, u)$ by multiplication by the matrix $\mathbf{t}$. Introduce the notation $\mathbf{b}(u, u)$ to denote the portion of $\mathbf{corr}(u, u)$ extending from $-N$ to $N$. So what we are saying is that if $u$ is a function such that $\mathbf{corr}(u, u)$ has support in $[-N, N]$, then $Cu$ has this same property and

$$\mathbf{b}(Cu, Cu) = \mathbf{t}\mathbf{b}(u, u).$$

Suppose that 1 is a simple eigenvalue of $\mathbf{t}$ and that all other eigenvalues satisfy $|\lambda| < 1$. Then if $\mathbf{v}$ is a vector such that $\mathbf{1} \cdot \mathbf{v} = 1$, where $\mathbf{1}$ denotes the row vector of all ones, then

$$\mathbf{1} \cdot T^n \mathbf{v} = 1$$

and $T^n\mathbf{v}$ converges (geometrically) to the unique eigenvector $\mathbf{e}$ of $\mathbf{T}$ with eigenvalue 1 and satisfying $\mathbf{1} \cdot \mathbf{e} = 1$.

So suppose we start with an initial function $u_0$ satisfying $\mathbf{corr}(u_0, u_0) = \delta_0$. Then the

$$\mathbf{b}_i := \mathbf{b}(C^i u_0, C^i u_0)$$

are converging to the eigenvector $\mathbf{e}$. Suppose we set

$$u_i := C^i u_0.$$

Then we see that the $\|u_i\|$ converge, so the $u_i$ form a weakly compact subset of $L_2$, hence we can choose a weakly convergent subsequence in the $L_2$ sense, which proves that the limit $\phi$ is in $L_2$. Furthermore, condition (3.7) translates into the condition that the central column of $\mathbf{T}$ and hence of $\mathbf{t}$ is $\delta_0$, so this is an, and hence the, eigenvector $\mathbf{e}$. This implies that the limit $\phi$ is orthogonal to its translates. To summarize, we have proved the following theorem of Lawton

**Theorem 3.2.1** *If (3.7) holds and* 1 *is a simple eigenvalue of* $\mathbf{t}$ *(with the other eigenvalues less than one in absolute value) then the infinite product converges to an element of* $L_2$ *which is orthogonal to its translates by integers.*

## 3.3 Strang's theorem.

We will prove a theorem of Strang which says that starting with the initial function $\phi^0 = \mathbf{1}_{[0,1]}$ and assuming the Lawton condition that $\mathbf{t}$ has 1 as an isolated eigenvalue, with all the other eigenvalues strictly less than one in absolute value, then the sequence $\phi^i = C^i\phi^0$ converges in $L_2$ to the scaling function. For simplicity, let us assume that our vector $h$ is supported on $[0, N]$ as assumed above and with $h(0) \neq 0$ and $h(N) \neq 0$.

**Lemma 3.3.1** *The functions* $\phi^i$ *are all supported on* $[0, N]$.

**Proof by induction.** This is clearly true for $\phi^0$ which is supported on $[0, 1]$. So assume that it is true for $\phi^n$. The right hand side of the iterative definition

$$\phi^{n+1}(t) = \sum 2h(k)\phi^n(2t - k)$$

vanishes unless one of the terms

$$2t - N, 2t - N + 1, \ldots, 2t - 1, 2t$$

belongs to $[0, N]$. But this requires that $2t \geq 0$ and $2t - N \leq N$ which says that $t \in [0, N]$. QED

By passing to the limit we know that supp $\phi \subset [0, N]$. We will discuss more detailed information about supp $\phi$ later on in this section. In any event, the vectors

$$\mathbf{corr}(\phi, \phi), \quad \mathbf{corr}(\phi, \phi^i) \quad \text{and} \quad \mathbf{corr}(\phi^i, \phi^i)$$

are all supported in $[-N, N]$. We also recall (2.17) which says

$$\sum_k \phi(x + k) \equiv 1 \tag{3.16}$$

in one dimension. Let

$$\mathbf{a} := \mathbf{corr}(\phi, \phi), \quad \mathbf{a}^i := \mathbf{corr}(\phi^i, \phi^i) \quad \text{and} \quad \mathbf{b}^i := \mathbf{corr}(\phi^i, \phi)$$

so that

$$\mathbf{a} = \mathbf{ta}, \quad \mathbf{a}^{i+1} = \mathbf{ta}^i, \quad \text{and } \mathbf{b}^{i+1} = \mathbf{tb}^i.$$

Let $\mathbf{1}$ denotes the row vector consisting of all ones, and let $\delta$ denote the column vector with $\delta(0) = 1$ and all other entries zero so that $\delta$ is the unique eigenvector of $\mathbf{t}$ with eigenvalue 1 satisfying $\mathbf{1} \cdot \delta = 1$. Then we have

$$\mathbf{a}^0 \delta,$$

since $\phi^0 = \mathbf{1}_{[0,1]}$ and hence $\mathbf{a}^i = \delta$ for all $i$. From the fact that $T\phi = \phi$ we know that $\mathbf{a}$ is an eigenvector of $\mathbf{t}$, hence some multiple of $\delta$. Hence all the scalar products $(\phi, \phi_k)$ vanish for $k \neq 0$, and $||phi||^2 = c \neq 0$. By Plancherel we may write

$$||\phi|^2 = ||\hat{\phi}||^2 = \sum_\ell \int_0^{2\pi} |\hat{\phi}(\xi + 2\pi\ell)|^2 d\xi$$

and so by Fubini

$$||phi||^2 = \int_0^{2\pi} \Phi(\xi) d\xi$$

where

$$\Phi(\xi) := sum_\ell |\hat{\phi}(\xi + 2\pi\ell)|^2.$$

But

$$c\delta(k) = (\phi, \phi_k) = \phi \star \tilde{\phi}(k) = \int |\hat{\phi}(\xi)|^2 e^{ik\xi} d\xi = \int_0^{2\pi} \Phi(\xi) e^{ik\xi} d\xi.$$

This means that all the Fourier coefficients of $\Phi$ vanish except the constant term and hence

$$\Phi(\xi) \equiv \frac{c}{2\pi}.$$

From the scaling equation we know that if we factor $\ell$ as a power of two times an odd number, $\ell = 2^r(2n + 1)$, then

$$\hat{\phi}(2\pi\ell) = \prod_{j=1}^{r-1} m(2^{r-j}(2n + 1)\pi) H(2n + 1)\pi)\hat{\phi}((2n + 1)\pi) = 0.$$

So $c = 2\pi\Phi(0) = 2\pi|\hat{\phi}(0)|^2 = 1$. In other words

$$\mathbf{a} = 1.$$

This is just another proof of Lawson's theorem.

Now

$$\mathbf{1} \cdot \mathbf{b}^0 = \sum_k \int_0^1 \overline{\phi(x-k)} dx = 1$$

by (3.16). Hence $\mathbf{1} \cdot \mathbf{b}^i = 1$ for all $i$, and therefore the $\mathbf{b}^i$ converge to $\delta$. In particular, the zero component of the $\mathbf{b}^i$, which is just $(\phi^i, \phi)$ converges to 1. Therefore

$$\|\phi^i - \phi\|^2 = (\phi^i, \phi^i) - 2\text{Re } (\phi^i, \phi) + (\phi, \phi) \to 0.$$

QED

We conclude with some comments about the support of $\phi$. We know it is supported in $[0, n]$. We claim that it is not supported in any subinterval. Indeed, let $a = \min x,\ x \in \text{supp}(\phi)$ and $b = \max x\ x \in \text{supp}(phi)$. The function $\phi(2t - k)$ is supported in the interval

$$\left[\frac{a+k}{2}, \frac{b+k}{2}\right]$$

so from the scaling equation we conclude that

$$[a, b] = \left[\frac{a}{2}, \frac{b+N}{2}\right]$$

and hence

$$a = 0, \qquad b = N.$$

# Chapter 4

# Algebraic constructions.

We look for trigonometric polynomials $H$ with real coefficients which satisfy our conditions

$$|H(\xi)|^2 + |H(\xi+\pi)|^2 \equiv 1 \tag{4.1}$$

and

$$H(0) = 1 \tag{4.2}$$

so

$$H(\pi) = 0. \tag{4.3}$$

and vanish to a high order at $\pi$. So we write

$$H(\xi) = \left(\frac{1+e^{-i\xi}}{2}\right)^n B(\xi)$$

and set

$$P(\xi) := |H(\xi)|^2 = H(\xi)H(-\xi).$$

Since

$$\left(\frac{1+e^{-i\xi}}{2}\right)^n \left(\frac{1+e^{-i\xi}}{2}\right)^n = \left(\frac{e^{i\xi/2}+e^{-i\xi/2}}{2}\right)^{2n} = \left(\cos^2\frac{\xi}{2}\right)^n$$

we have

$$P(\xi) = \left(\cos^2\frac{\xi}{2}\right)^n A(\xi)$$

where

$$A(\xi) = B(\xi)B(-\xi) := \tilde{P}(\cos\xi).$$

Set

$$y := \sin^2\xi$$

so we can write

$$\tilde{P}(\cos\xi) = \tilde{P}(1-2y) =: Q(y).$$

In short

$$P(\xi) = (1-y)^n Q(y).$$

Also,

$$A(\xi+\pi) = \tilde{P}(-\cos\xi) = \tilde{P}(2y-1) = \tilde{P}(1-2(1-y)) = Q(1-y)$$

so the key equation

$$P(\xi) + P(\xi+\pi) \equiv 1$$

becomes

$$(1-y)^n Q(y) + y^n Q(1-y) \equiv 1. \tag{4.4}$$

Our problem is to solve this equation and then factor the corresponding $P$ as $H(\xi)H(-\xi)$.
sectionFinding $Q$ The theorem on partial fractions says that there are uniquely determined coefficients $C_k, C'_k$ such that

$$\frac{1}{y^n(1-y)^n} = \sum \frac{C_k}{y^k} + \sum \frac{C'_k}{(1-y)^k}$$

and symmetry demands that $C_k = C'_k$. Clearing denominators shows that there is a unique polynomial $Q_n$ of degree $\leq n-1$ solving (4.4). Explicitly, since any solution of (4.4) satisfies

$$Q(y) = (1-y)^{-n}(1-y^n Q(1-y))$$

we can find $Q_n$ by simply ignoring the coefficients of degree higher than $n-1$ in the power series expansion of $(1-y)^{-n}$. so by the binomial formula

$$Q_n(y) = 1+ny+\frac{n(n+1)}{2}y^2+\cdots+\binom{2n-2}{n-1}y^{n-1} = (1-y)^{-n}+o(y^n). \tag{4.5}$$

Notice that

$$(1-y)^n Q_n(y) = (1-y)^n[(1-y)^{-n}+o(y^n)] = 1+o(y^n).$$

Thus $(1-y)^n Q_n(y)$ is a polynomial of degree $2n-1$ which takes the value 1 at zero, and whose first $n-1$ derivatives vanish at 0 and which vanishes together with its first $n-1$ derivatives at $y=1$. For the low values $n = 1, 2, 3$ we have the following table

| $n$ | $Q_n(y)$ | $(1-y)^n Q_n(y)$ |
|---|---|---|
| 1 | $1$ | $1-y$ |
| 2 | $1+2y$ | $1-3y^2+2y^3$ |
| 3 | $1+3y+6y^2$ | $1-10y^3+15y^4-6y^5$ |

.

## 4.1 The Riesz theorem and the Daubechies polynomials.

So, for a "minimal solution" we have

$$P(\xi)=\left(\cos^2\frac{\xi}{2}\right)^n Q_n\left(\sin^2\frac{\xi}{2}\right).$$

Going back to the notation

$$A(\xi)=Q_n(\sin^2\frac{\xi}{2})=\tilde{P}(\cos\xi)$$

we need to factor this as

$$A(\xi)=B(\xi)B(-\xi)$$

which is possible due to a theorem of Riesz:

**Theorem 4.1.1** **[Riesz]** *If*

$$A(\xi)=\sum_{k=0}^{m}a_k\cos^k\xi$$

*is a polynomial with real coefficients (with $a_m\neq 0$) such that*

$$A(\xi)\geq 0$$

*for all real $\xi$ then there is a trigonometric polynomial*

$$B(\xi)=\sum_{k=0}^{m}b_ke^{-ik\xi}$$

*such that*

$$A(\xi)=B(\xi)B(-\xi).$$

*If $A(0)=1$ we may choose $B$ so that $B(0)=1$.*

**Proof.** Factor

$$A(\xi)=a_m\prod(\cos\xi-c_j)$$

where the $c_j$ are real or occur in complex conjugate pairs. Set

$$z:=e^{-i\xi}.$$

Then

$$A(\xi)=a_m\prod\left(\frac{z+z^{-1}}{2}-c_j\right).$$

We use the identity

$$\frac{z+z^{-1}}{2}-\frac{s+s^{-1}}{2}=-\frac{1}{2s}(z-s)(z^{-1}-s). \tag{4.6}$$

- If $c_j \in \mathbf{R}$ and $|c_j| \geq 1$ then there exists $s \in \mathbf{R}$ such that
$$c_j = \frac{s + s^{-1}}{2}.$$
Then we get the factorization
$$\frac{z + z^{-1}}{2} - c_j = -\frac{1}{2s}(z - s)(z^{-1} - s).$$

- If $c_j \in \mathbf{R}$ and $|c_j| < 1$ then $c_j = \cos\alpha$ so if we set $s = e^{i\alpha} \neq \pm 1$ the factorization of $A$ above contains the term $(\cos\xi - \cos\alpha)$ which can changes sign near $\xi = \alpha$ and so must occur an even number of times. It thus factors.

- In the complex case we can solve for $s$ i
$$c_j = \frac{s + s^{-1}}{2}$$
and then get the corresponding equation for $\overline{c_j}$ by replacing $s$ by $\overline{s}$. Then
$$\left(\frac{z + z^{-1}}{2} - c_j\right)\left(\frac{z + z^{-1}}{2} - \overline{c_j}\right) = \left(\frac{z + z^{-1}}{2} - \frac{s + s^{-1}}{2}\right)\left(\frac{z + z^{-1}}{2} - \frac{\overline{s} + \overline{s}^{-1}}{2}\right)$$
$$= \frac{1}{4|s|^2}(z - s)(z^{-1} - s)(z - \overline{s})(z^{-1} - \overline{s})$$
$$= \frac{1}{4|s|^2}(z^2 - 2\mathrm{Re}\; sz + |s|^2)(z^{-2} - 2\mathrm{Re}\; sz^{-1} + |s|^2)$$
which is again a factorization of the desired type. QED