Fourier Analysis, Multiresolution Analysis and Dilation Equations

David Malone

September 1997

Declaration

I do declare that:

- This wasn't anyone's homework.
- As work goes it's mine, mine, all mine, [*] and where I've used other peoples work I've said so. [†]
- You know what the library can do with this when they get it? More or less whatever they want. They should keep my name on it though.

Or — at your discretion:

- This has not previously been submitted as an exercise for a degree at this or any other University.
- This work is my own except where noted in the text.
- The library should lend or copy this work upon request.

David Malone (August 4, 1998).

[*]Keep your greasy paws off it.
[†]So there!

Summary

This thesis has essentially two parts.

The first two chapters are an introduction to the related areas of Fourier analysis, multiresolution analysis and wavelets. Dilation equations arise in the context of multiresolution analysis. The mathematics of these two chapters is informal, and is intended to provide a feeling for the general subject. This work is loosely based on two talks which I gave, one during the 1997 Inter-varsity Mathematics competition and the other at the 1997 Dublin Institute for Advanced Studies Easter Symposium.

The second part, Chapters 3 and 4, contain original work. Chapter 3 provides a new formal construction of the Fourier transform on $L^p(\mathbb{R}^n)$ $(1 \leq p \leq 2)$ based on the ideas introduced in Chapter 2.

The idea is to take some basic properties of the Fourier transform and show we can construct a bounded operator on $L^2(\mathbb{R})$ with these properties. I do this by constructing an operator on each level of the Haar multiresolution analysis, which I then show is well enough behaved to be extended by a limiting process to all of $L^2(\mathbb{R})$.

Some of the important properties of the Fourier transform are also derived in terms of this construction, and the generalisations to $L^p(\mathbb{R}^n)$ are explored.

Chapter 4 builds on the work of Chapters 3 and provides a uniqueness result for the Fourier transform. While searching for this result I also establish a related result for dilation equations (a subject also introduced in Chapter 2).

Here the exact set of properties which were used to define the Fourier transform are varied in an effort to discover which are merely consistent with the Fourier transform and which strong enough to define it. I end up examining sets of dilation equations and determining when these will have a unique solution.

Acknowledgments

I'm not really sure who I should acknowledge. When listing people you are sure to miss people to whom credit is due and perhaps even accidently include people who have had no real input. For instance:

I'd like to thank Elvis for all those lunchtime chats in which he tried to explain why the "admissibility condition" wasn't a terrible name. Also the comments Jimmi Hendrix and Jesus made on layout were very helpful.

However, everyone seems to do their best when compiling acknowledgments, and why should I be the exception?

I should first thank my supervisor, Richard Timoney, and other members of the School of Mathematics who have provided me with access to a far better source of information than books — experience. That is not to say they did not provide me with access to books too, the library in the basement is quite comprehensively stocked. Sarah Ziesler also performed a broad assessment of and commentary on the second half of this work — always a useful thing.

My family and the Trinity Foundation have provided support of a very practical nature. My family house, feed and put up with me. The Trinity Foundation have funded my time as a postgraduate. The secretarial staff of the School of Mathematics have always provided help with photocopying, stapling and looking over printouts before I use them.

John Lewis of DIAS and the committee of the Dublin University Mathematics Society provided me with opportunities to address two quite different audiences. Both were educational and enjoyable experiences.

There is a long list of people who have helped by talking through things with me, volunteering for proof reading duty or just listening to me babble about whatever I was thinking about at the time. This list of people includes, but is not limited to: Peter Clifford, Shane Crowe, Ian Dowse, Ken Duffy, Eoin Hickey, Lesley Malone, Sharon Murphy, Ray Russell, Karl Stanley and John Walsh.

Finally, Tim Murphy may pretend otherwise but does actually know more about LaTeX than anyone else. An invaluable mentor for all those using TeX.

Contents

1 Introducing Fourier Analysis **1**

1.1 Introduction . . . 1

1.2 Fourier Series and Integrals . . . 1

1.3 L^p and Convergence . . . 4

1.4 Why bother with Fourier Analysis? . . . 6

1.5 The Windowed Fourier Transform . . . 9

1.6 The Continuous Wavelet Transform . . . 10

1.7 Conclusion . . . 13

2 An Introduction to Multiresolution Analysis **15**

2.1 Introduction . . . 15

2.2 What is MRA? . . . 15

2.3 Three Examples . . . 18

2.3.1 The Haar MRA . . . 18

2.3.2 Band-Limited MRA . . . 20

2.3.3 Daubechies' generating function . . . 24

2.4 Dilation Equations . . . 25

2.4.1 Integrability . . . 25

2.4.2 Orthonormality . . . 26

2.4.3 Ability to Approximate . . . 26

2.5 Discrete Wavelets . . . 27

2.6 Back to the examples . . . 29

2.6.1 Haar Wavelets . . . 30

2.6.2 Shannon Wavelets . . . 31

2.6.3 Daubechies' Wavelets . . . 32

2.7 How to draw these beasties . . . 32

2.8 Conclusion 34

3 A New Construction for the Fourier Transform 36

3.1 Introduction 36

3.2 Defining $\mathcal{F}$ on our MRA 37

3.3 Extending $\mathcal{F}$ to $L^2(\mathbb{R})$ 42

3.4 Back to the traditional 45

3.5 Can we do better than that? 48

3.6 Extending properties of $\mathcal{F}$ 50

3.7 The inverse Fourier transform 52

3.8 Higher dimensions 53

3.9 $L^p(\mathbb{R}^n)$ for other p. 57

3.10 Conclusions 61

4 A Uniqueness Result for the Fourier Transform 64

4.1 Introduction 64

4.2 A simple start 65

4.3 A useful counterexample 67

4.4 How many solutions? 70

4.5 Some more dilation equations 74

4.6 Pinning it down 77

4.7 Conclusion 81

5 Future Work 82

List of Figures

1.1 Averages of $\cos nx$ . . . 2
1.2 $e^{-x^2}\cos(7x)$ and its Fourier transform . . . 7
1.3 5 notes and its Fourier transform . . . 8
1.4 Possible windows for the WFT . . . 9
1.5 FT and WFT of 5 notes. . . . 11
1.6 A wavelet and its Fourier Transform . . . 12

2.1 The approximation process. . . . 16
2.2 $g(x)$ for the Haar MRA. . . . 19
2.3 Improving Approximations. . . . 21
2.4 $g(x) = \frac{\sin \pi x}{\pi x}$. . . . 22
2.5 $g(x)$ for one of Daubechies' MRAs. . . . 24
2.6 $w(x)$ for the Haar MRA. . . . 30
2.7 $w(x)$ for the Band-Limited MRA. . . . 31
2.8 $w(x)$ for one of Daubechies MRA's. . . . 32
2.9 C code for estimating g . . . 34

3.1 $\left|\frac{1-e^{-i\omega}}{i\omega}\right|$ . . . 59

4.1 $\phi(x)$ . . . 68

Chapter 1

Introducing Fourier Analysis

1.1 Introduction

Fourier analysis has become an extremely useful subject in mathematics, science and engineering. This section will explain the idea behind the Fourier transform and also show how this idea leads to wavelet analysis. The emphasis here is on an intuitive understanding and motivation for the steps taken, not a formal justification — which can be read in any mathematical book on these subjects [15, 16]. The actual process of calculating and using Fourier analysis is dealt with in most engineering mathematics books, for example [10].

The basic idea of Fourier analysis is: Suppose we have some function f, which we know is a sum of terms something like $a_n \cos(nx)$, but we don't know what the a_n are. Then how do we find exactly what these terms are? For engineers this amounts to looking at how much (a_n) of what basic frequencies (n) make up a given signal (f).

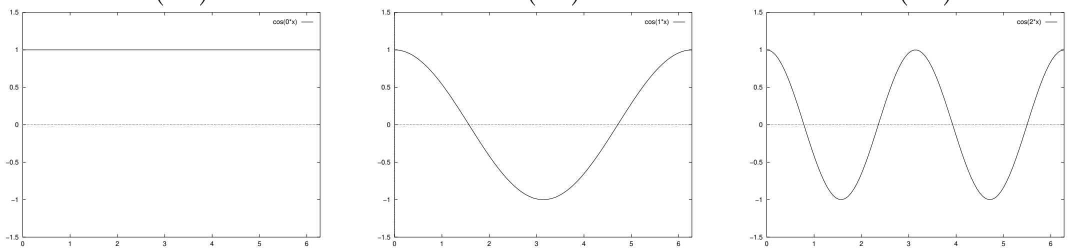
1.2 Fourier Series and Integrals

To keep the details simple we will start by looking at a 'Fourier cosine series'. Suppose we have some function f which we know is a sum* of the form:

$$f(x) = \sum_{n=0} a_n \cos nx,$$

*We have deliberately left the upper limit of the sum out, as we do not want to look at the issue of convergence yet.

$\cos(0x)$ $\cos(1x)$ $\cos(2x)$



Average 1 Average 0 Average 0

Figure 1.1: Averages of $\cos nx$

but we do not know what the exact values of the a_n are. How do we go about finding what these a_n are? There are essentially three tricks to finding the a_n. First, $f(x)$ must have period which divides 2π, because all the $\cos nx$ repeat themselves after 2π. The second trick involves looking at the average of $\cos nx$ on $[0, 2\pi]$. Looking at Figure 1.1 we see that $\cos nx$ has average 0 whenever n is not 0. We can exploit this by averaging f.

$$\int_0^{2\pi} f(x)\,dx = \sum_{n=0} a_n \int_0^{2\pi} \cos nx\,dx = 2\pi a_0$$

So now we have found a_0. We would like to be able to use the same trick to pick out the rest of the a_n. This is where the third trick comes in. We look at what happens when we multiply $\cos nx$ by $\cos mx$:

$$\cos nx \cos mx = 1/2\,(\cos(m+n)x + \cos(m-n)x)$$

Averaging both sides (again over $[0, 2\pi]$) we see that we get a contribution of $1/2$ if $m+n = 0$ and another $1/2$ if $n-m = 0$. If we get both contributions then $n = m = 0$, which we have dealt with. Otherwise we need $n = m$ to get a contribution (because we are only worried about $n \geq 0$ at the moment).

Using what we have just learned we try multiplying f by $\cos mx$ before averaging.

$$\begin{aligned}\int_0^{2\pi} f(x)\cos mx\,dx &= \sum_{n=0} a_n \int_0^{2\pi} \cos nx \cos mx\,dx \\ &= \pi a_m\end{aligned}$$

So now, all formalities out of the way we have found a way to determine the a_n given the function f.

Further examination of the usual trigonometric identities lets us do some variations on this theme. By using the formulas:

$$\cos nx \sin mx = 1/2\,(\sin(n+m)x - \sin(n-m)x)$$

and

$$\sin nx \sin mx = 1/2\,(\cos(n-m)x - \cos(n+m)x)$$

and also remembering that $\sin nx$ averages to zero (over $[0, 2\pi]$) regardless of the value of n we may deal with a more general situation. This time suppose that f is expressible in the following form:

$$f(x) = \sum_{n=0} a_n \cos nx + b_n \sin nx.$$

By going through the process of multiplying by $\cos mx$ or $\sin mx$ and averaging we eventually get expressions for a_n and b_n.

$$\begin{aligned}
a_0 &= \frac{1}{2\pi}\int_0^{2\pi} f(x)\,dx \\
b_0 &= 0 \qquad \text{(doesn't matter)} \\
a_n &= \frac{1}{\pi}\int_0^{2\pi} f(x)\cos mx\,dx \\
a_n &= \frac{1}{\pi}\int_0^{2\pi} f(x)\sin mx\,dx
\end{aligned}$$

Naturally we can make many variants of this. The a_n and b_n could be objects in some real vector space, we could work on $[0, 1]$ instead of $[0, 2\pi]$ or we could just use sin. The two most important variations are replacing trigonometric functions with e^{inx} and replacing sums with integrals.

Replacing $\sin x$ and $\cos x$ with e^{ix} is generally viewed as a simplification. We retain the same degree of generality (we now work with a sum from $-\infty$ to ∞) but we only need one formula. If we suppose that:

$$f(x) = \sum_{n=-\infty}^{\infty} c_n e^{inx}$$

and remember that e^{inx} has period dividing 2π, we might be tempted to average over $[0, 2\pi]$ again. The average of e^{inx} is zero over the range $[0, 2\pi]$ — unless n is 0 when the average is 1. This combined with the fact:

$$e^{inx}e^{imx} = e^{i(n+m)x},$$

means that $f(x)e^{-imx}$ has mean c_m over $[0, 2\pi]$. So we get a simple expression for the c_m:

$$c_m = \frac{1}{2\pi}\int_0^{2\pi} f(x)e^{-imx}\,dx.$$

Now we attempt to replace our sums with integrals. This requires a certain extra leap of faith as regards convergence. If we believe that in some sense $e^{i\omega x}$ has average 0 for all $\omega \in \mathbb{R} \setminus \{0\}$ then we can hope that if:

$$f(x) = \int_{-\infty}^{\infty} c(\omega)e^{i\omega x}\,d\omega,$$

then there is some chance that:

$$c(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} f(x)e^{-i\omega x}\,dx.$$

In this case $c(\omega)$ is called the *Fourier transform* of f. As we can see there is a certain degree of symmetry in the expressions for f and c. For this reason people often move around factors of 2π or $\sqrt{2\pi}$ to try to make the situation even more symmetric.

The next two questions that arise are: what functions can we write in these forms, and where don't we have to worry about convergence problems? The answers to these questions are related, and the relation is linked to the symmetry we have just noted.

1.3 L^p and Convergence

When looking at the convergence of something like:

$$\int_{-\infty}^{\infty} f(x)e^{-i\omega x}\,dx,$$

for various values of ω the first thing we can note is that $|e^{-i\omega x}| = 1$, so all we really need to worry about is:

$$\int_{-\infty}^{\infty} f(x)\, dx.$$

However, we may as well just look at $|f|$ because it might turn out that f was strictly positive, or that multiplying f by $e^{-i\omega x}$ made it strictly positive. This leaves us looking at:

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

This is called the L^1 norm of f and is usually written $\|f\|_1$. On the set of f where $\|f\|_1$ is finite (this is called $L^1(\mathbb{R})$), this $\|.\|_1$ is a norm in the usual *normed vector space* sense — barring some complications regarding equivalence classes.

On this space, $L^1(\mathbb{R})$, we find the Fourier transform is quite well behaved. It is reasonably easy to show that if f is in $L^1(\mathbb{R})$ then its Fourier transform is a continuous and bounded function of ω. However this doesn't really reflect the symmetry we noticed. We started with $L^1(\mathbb{R})$, applied the Fourier transform and got continuous and bounded functions. The symmetry would suggest that we search for a space which the Fourier transform sends to itself. Here, perhaps, the Fourier transform might even be invertible.

A similar norm to $\|.\|_1$, but one closer to the traditional Euclidean norm would be:

$$\|f\|_2 = \sqrt{\int_{-\infty}^{\infty} |f(x)|^2\, dx}.$$

We call the space of functions where this new norm is finite $L^2(\mathbb{R})$. Again, ignoring some complications with equivalence classes of functions, this is also a normed vector space. It also turns out to be the ideal space for Fourier analysis. It can be shown (though it takes some work) that the Fourier transform is a linear, continuous and invertible transform, both to and from this normed vector space: $L^2(\mathbb{R})$.

When looking at Fourier series and other problems we can generalise our definition of spaces of this sort. A few examples best illustrate the idea.

$$\begin{aligned}
L^1(\mathbb{R}) &= \left\{ f : \mathbb{R} \to \mathbb{C} \middle| \int_{-\infty}^{\infty} |f(x)|\, dx < \infty \right\} \\
L^2([0, 2\pi]) &= \left\{ f : [0, 2\pi] \to \mathbb{C} \middle| \int_{0}^{2\pi} |f(x)|^2\, dx < \infty \right\} \\
L^3(\mathbb{N}) &= \left\{ a_n \in \mathbb{C} \middle| n \in \mathbb{N}, \sum_{n=0}^{\infty} |a_n|^3 < \infty \right\}
\end{aligned}$$

The first example we have already seen. The second relates to the Fourier series for 2π periodic functions. The last gives an example of other sorts of L^p type spaces which we can consider. In general for X a set[†] with a positive measure μ we could define $L^p(X, \mu)$ to be the set:

$$\left\{ f : X \to \mathbb{C} \middle| f \text{ measurable and } \int_X |f|^p\, d\mu < \infty \right\}$$

In this context our third example was no more than this general definition on $\mathbb{N}$ with the counting measure. This family $L^p(\mathbb{N}, \text{counting})$ is often just written l^p, as they are rather commonly encountered spaces.

1.4 Why bother with Fourier Analysis?

Fourier analysis (which is this process of writing things in terms of e^{inx}), has some good reasons for being of interest. Naturally all these reasons are interlinked. We'll start from a mathematical reason and then work toward a practical reason.

One of the important topics which mathematicians deal with is the study of differential equations. If we look at $v(x) = e^{i\omega x}$, then we see that v is an *eigenvector* for differentiation. That is:

$$\frac{d}{dx} v = \frac{d}{dx}\left(e^{i\omega x}\right) = i\omega e^{i\omega x} = i\omega v.$$

So differentiating v has a very simple effect on it (it gets multiplied by $i\omega$). If we are looking for a solution of some differential equation and we know the solution may be expressed as the sum (or integral) of terms like $e^{i\omega x}$ then, the differential equation should have a simple

[†]More technically, we need a measure space rather than just a set.

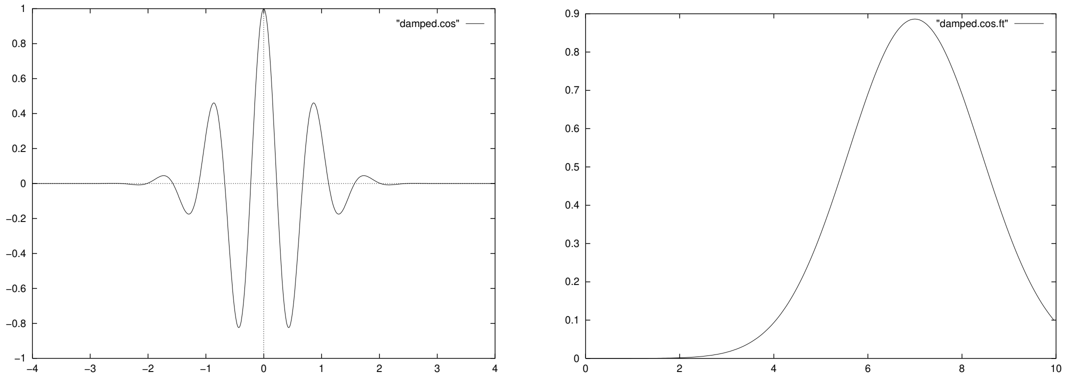


Figure 1.2: $e^{-x^2}\cos(7x)$ and its Fourier transform

form for each of these terms, which will hopefully be easy to solve. This means that Fourier analysis is likely to be useful for solving physical problems, which often involve differential equations.

An important differential equation which lends itself to this sort of Fourier analysis perfectly is the wave equation.

$$\frac{\partial^2}{\partial x^2} - \frac{1}{c^2}\frac{\partial^2}{\partial t^2} = 0$$

This equation describes many things, including the transmission of electromagnetic waves (light) and sound. For this reason Fourier analysis is of great interest to engineers working with light or sound signals. They think of the Fourier transform of a signal as showing "how much" of each frequency is in a signal. Also, in this analogy, the $\|.\|_2^2$ corresponds to the energy of the signal, making results easy to interpret.

An example may make this clearer. In Figure 1.2 we see a signal which we would expect to be mostly composed of a wave of angular frequency 7 (as it is mainly $\cos 7x$). If we look at its Fourier transform we see that there is a large peak near 7, as we would have hoped.

A more complicated example reveals both the strength and weakness of this sort of analysis. Suppose we have a signal[‡], which is someone playing a few notes. We could use our Fourier analysis to look for peaks to find which notes have been played. For instance

[‡]The signal in the case of sound could be considered to be the change in pressure of the air in the ear with respect to time.

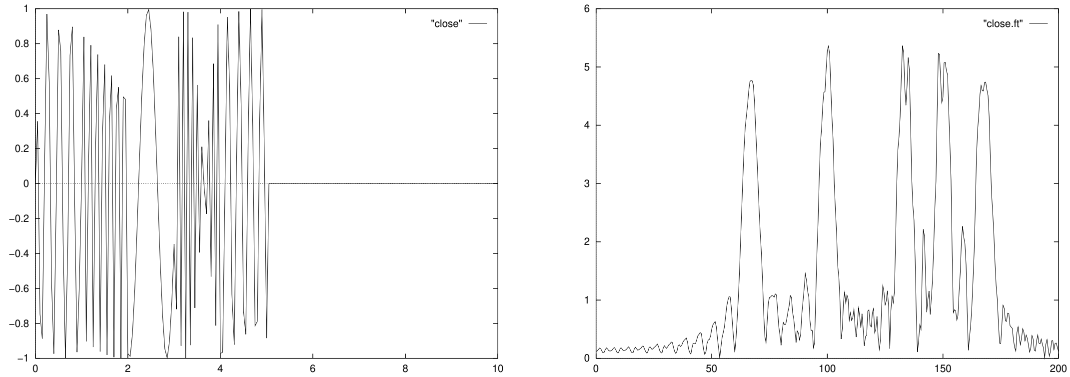
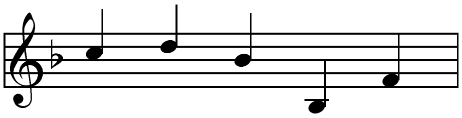


Figure 1.3: 5 notes and its Fourier transform

a signal[§] representing the following notes is shown in Figure 1.3, along with its Fourier transform.



In the Fourier transform we can quite clearly see the 5 peaks, one for each note, and see that each peak is about the same size, and so corresponds to roughly the same total energy. A problem arises when we try to get some indication about when each note was played. The information must be there — as the Fourier transform is invertible — but it is not easily accessible. For all we know we could be looking at the Fourier transform of the following instead[¶].

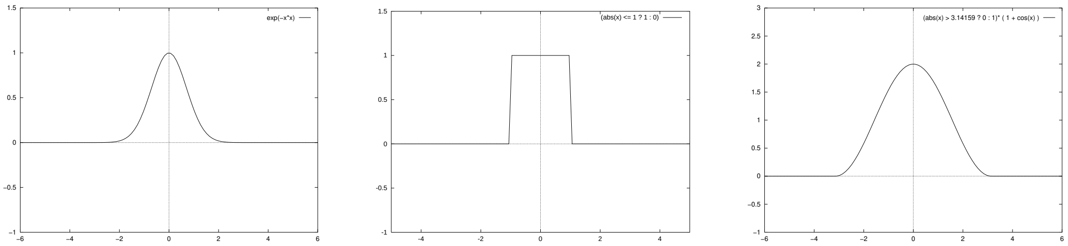
[§]This set of notes shown could be represented by:

$$p(t) = \begin{cases} \cos(100\alpha^{7}t) & 0 \le t < 1 \\ \cos(100\alpha^{9}t) & 1 \le t < 2 \\ \cos(100\alpha^{5}t) & 2 \le t < 3 \\ \cos(100\alpha^{-7}t) & 3 \le t < 4 \\ \cos(100\alpha^{0}t) & 4 \le t < 5 \\ 0 & \text{otherwise} \end{cases},$$

where $\alpha = \sqrt[12]{2}$ is the ratio of the frequencies of adjacent semitones in the modern tempered scale. A good introduction to the physics and mathematics of music is [7].

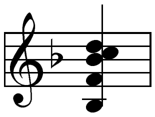
[¶]In this case the signal might be:

$$p(t) = \begin{cases} \cos(100\alpha^{7}t) + \cos(100\alpha^{9}t) + \cos(100\alpha^{5}t) + \cos(100\alpha^{-7}t) + \cos(100\alpha^{0}t) & 0 \le t < 1 \\ 0 & \text{otherwise} \end{cases}.$$



e^{-x^2} $\chi_{[-1,1]}$ $(1+\cos x)\chi_{[-\pi,\pi]}$

Figure 1.4: Possible windows for the WFT



1.5 The Windowed Fourier Transform

A step toward getting more *localised* information about the frequencies in a signal is to examine the signal through a "window". The idea is quite simple — if we want to know what frequencies are present in some region of a signal, then we somehow cut out part of the signal around the region of interest and look at the Fourier transform of this new signal.

Some typical windows are shown in Figure 1.4. The way we use such a window is we center it over the part of the signal we are interested in, and then multiply point by point. Suppose our window is $b(x)$ and our signal is $f(x)$. Then we window our signal at position x_0 and get:

$$f(x)b(x-x_0).$$

Now we take the Fourier transform of this and, hopefully, get an idea of how much of what frequencies (ω) are present near x_0.

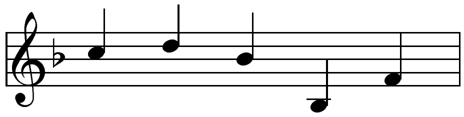
$$\breve{f}(x_0,\omega)=\int f(x)b(x-x_0)e^{-i\omega x}\,dx.$$

We are now getting into an area less strictly defined than that of the Fourier transform. This new "windowed Fourier transform" (WFT) clearly depends on our choice of the window b. Also, we now have a function of two variables, which provides us with more information, but is somehow more redundant. In fact there is also a way to get f back from its WFT — again not letting formalities get in the way:

$$f(x) = C \int \int e^{i\omega x} \overline{b(x - x_0)} \breve{f}(x_0, \omega) \, d\omega \, dx_0,$$

where C is inversely proportional to $\|b\|_2^2$ and $\overline{z}$ denotes complex conjugate of z.

This whole scheme is quite effective for frequency analysis. Going back to analysing



we now perform a WFT of the signal. Figure 1.5 shows both the Fourier transform and 5 slices of the windowed[‖] Fourier transform, centered at the time of the middle of each note played. We can see clear peaks in each of the these slices corresponding to the pitch of the note being played at that time. It would seem this method is a success, we can now tell when each frequency is contributing to the signal.

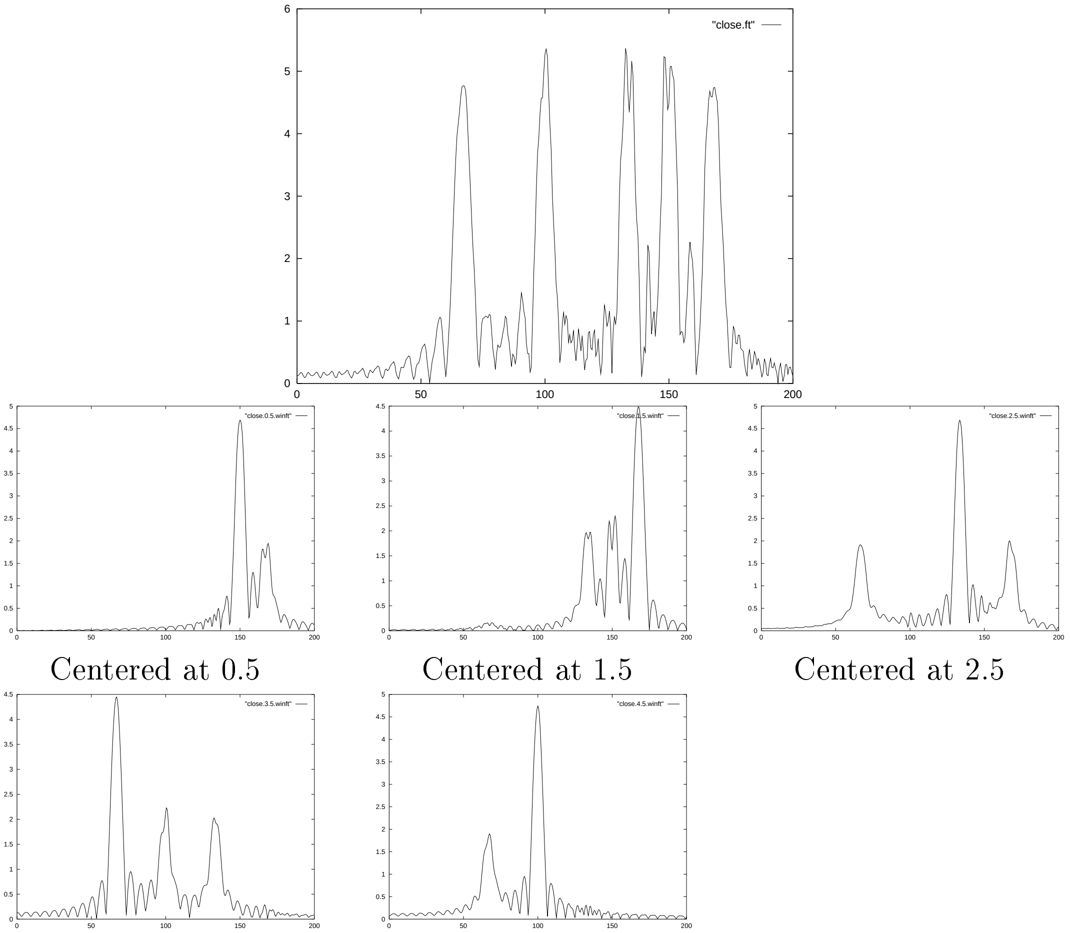
There is one drawback to this method though, which is related to choosing the window. The window is of fixed width, so we need to choose how wide to make it. This means we must either know the length of the "notes" of interest within the signal or we must experiment until we find a good width. If we make the window too narrow we will be unable to look at signals with a wavelength much longer than this width. If we make the window too wide we blur adjacent notes together.

For Figure 1.5 the window chosen had a width on the same scale as the length of the notes (we can actually see the adjacent notes as smaller peaks, so it may have been a little too wide). Trying to remove this limitation of the WFT leads us to the continuous wavelet transform.

1.6 The Continuous Wavelet Transform

The problem with the windowed Fourier transform is that the window is of fixed width. The correct direction to move in would seem to involve varying the width of the window. However if we just introduce a new parameter for window width, then we have a 3 parameter transform (ω for frequency, x_0 for position and say w for width), which some might view as getting a wee bit out of hand.

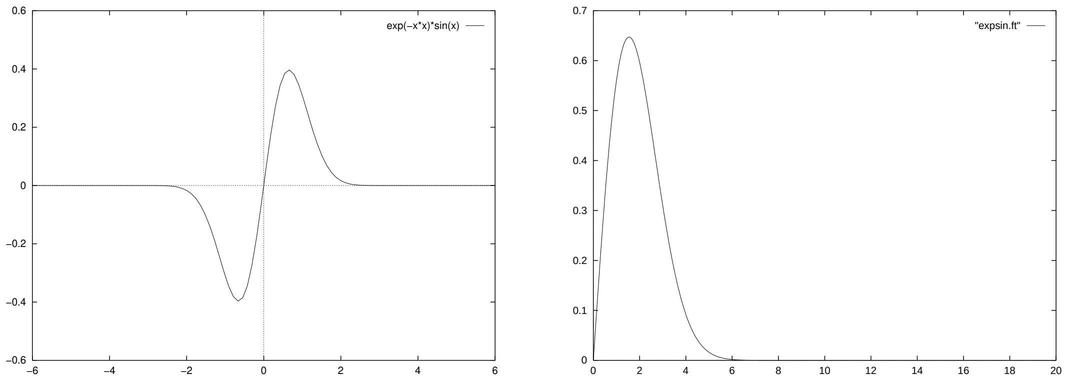
[‖]The window used was e^{-x^2}.



Centered at 0.5 Centered at 1.5 Centered at 2.5

Centered at 3.5 Centered at 4.5

Figure 1.5: FT and WFT of 5 notes.



ψ $\qquad$ $\mathcal{F}(\psi)$

Figure 1.6: A wavelet and its Fourier Transform

Realising that the width of the window can be related to the frequency which we are looking for is a useful trick. The shortest burst of some frequency that can be easily identified as being of that frequency is about half a wavelength** of said frequency. So if we choose a window width of about this length then we stand a good chance of picking up on that frequency.

So what we do is choose a window (say e^{-x^2}) and a wave (say $\sin x$), and glue them together to get a wavelet.

$$\psi(x) = e^{-x^2} \sin x$$

Now, in the same way as with the WFT, we center our wavelet over the signal.

$$\psi(x - x_0) = e^{-(x-x_0)^2} \sin(x - x_0)$$

We also note that if we want to search for frequency ω we just scale the whole thing by ω, as this produces a $\sin \omega x$ which will hopefully pick up on this frequence.

$$\psi\left(\omega(x - x_0)\right) = e^{-\omega^2(x-x_0)^2} \sin \omega(x - x_0)$$

This ψ is often called the "mother wavelet", and the changes in scale and position are often written as $\psi_{a,b}$ where:

$$\psi_{a,b}(x) = |a|^{-\frac{1}{2}} \psi\left(a^{-1}(x - b)\right).$$

**This can be made more formal, by Shannon's sampling theorem, see Theorem 2.1.

(note that $a = \omega^{-1}$, $b = x_0$ and the factor $|a|^{-\frac{1}{2}}$ is to normalise the wavelets in $L^2(\mathbb{R})$). With this notation the *continuous wavelet transform* (CWT) of some function/signal f is just:

$$F(a, b) = \int f(x)\overline{\psi_{a,b}(x)}\, dx.$$

Again there are reconstruction formulas, conditions[††] and even a "reciprocal" mother wavelet based on our choice of original mother wavelet ψ. The nitty gritty of this is spelt out in [8, 9].

All this freedom allowed by the choice of ψ explains much of the industry associated with wavelets. People choose mother wavelets to suit their own projects, which provides plenty of work for the numerical analysis of the following calculations. A lot of the buzz phrases associated with wavelets, such as "localised in time and frequency" are no more profound than can be seen in Figure 1.6 — where we note the wavelet is concentrated in one area, as is its Fourier transform .

There are also discrete wavelet transforms, similar to Fourier series, whose aim is to reduce the redundancy of this two parameter transform and give these wavelets useful properties such as orthogonality. These discrete wavelet transforms are largely built around a structure called multiresolution analysis (or multiscale analysis), a subject we will come back to in Chapter 2.

1.7 Conclusion

Here we have seen a little of Fourier analysis, a subject of interest to those involved with either pure or applied mathematics. We have dealt mostly with why it might be of interest to someone from a practical point of view, and said only little about the theory.

The most important result that we have missed is to do with relating the Fourier transform of a product of two functions to their individual Fourier transforms. This result involves the convolution of two functions:

$$(f * g)(x) = \int f(t)g(x - t)\, dt.$$

[††]It is interesting that the main condition required for the CWT is called the "admissibility" condition. Obviously someone used this phrase once and everyone else latched on to it. It should be recorded as one of the more vacuous phrases in mathematics, along with "normal" and "regular".

It states that the Fourier transform of a product is the convolution of the Fourier transforms (up to factors of $\sqrt{2\pi}$). Varying degrees of detail about this can be found in [15, 16, 10, 8].

On the practical side of things, the next most important result is probably the "Fast Fourier Transform", or FFT for short. This is a way of numerically calculating a Fourier transform from a set of sampled data far more quickly than just using numerical integration. This caused quite a stir when it came to light in the numerical world, as it changed an N^2 operation into a $N \log_2 N$ operation. Two whole chapters on implementation and usage of this can be found in [13].

On the theoretical side Fourier analysis has progressed in many directions. The definition of the Fourier transform can be extended to a large class of groups. Through this Fourier analysis has links with group representations. Many of the essays in [2] provide summaries of or introductions to these topics. There are also links with the theory of distributions and with the study of certain types of operators on spaces where we can use the Fourier transform. The Hilbert transform is probably the best known of these operators.

We have also looked at some practical variants of the Fourier transform, which has led us to the continuous wavelet transform. We will be moving on to look at how a stronger structure can be placed on these wavelet ideas to produce the discrete wavelet transform. These wavelet transforms have attracted similar attention to the attention received by the FFT when it appeared.

Chapter 2

An Introduction to Multiresolution Analysis

2.1 Introduction

In this section we will be introduced to some ideas relating to the discrete wavelet transform. The most important idea is that of a multiresolution analysis. Multiresolution analyses lead in a natural way to dilation equations — which feature strongly in many of the following sections. Also dilation equations lead us through the construction of the discrete wavelet transform.

2.2 What is MRA?

Multiresolution analysis could have several names applied to it, and indeed often does. The names multiscale analysis and multiresolution approximation are pseudonyms which provide further hints as to what it is all about. Maybe one of the more intuitive ways to approach MRA is from the approximation side.

When we want to approximate something we usually take several steps.

1. First we choose some function with which we will approximate. Maybe a spline of some sort, maybe something specially tailored for our approximation problem.

2. We translate our chosen function to various nodes, where each translated function will do its approximation. Often these nodes might be the integers.

3. We multiply each translated function by some carefully chosen coefficients.

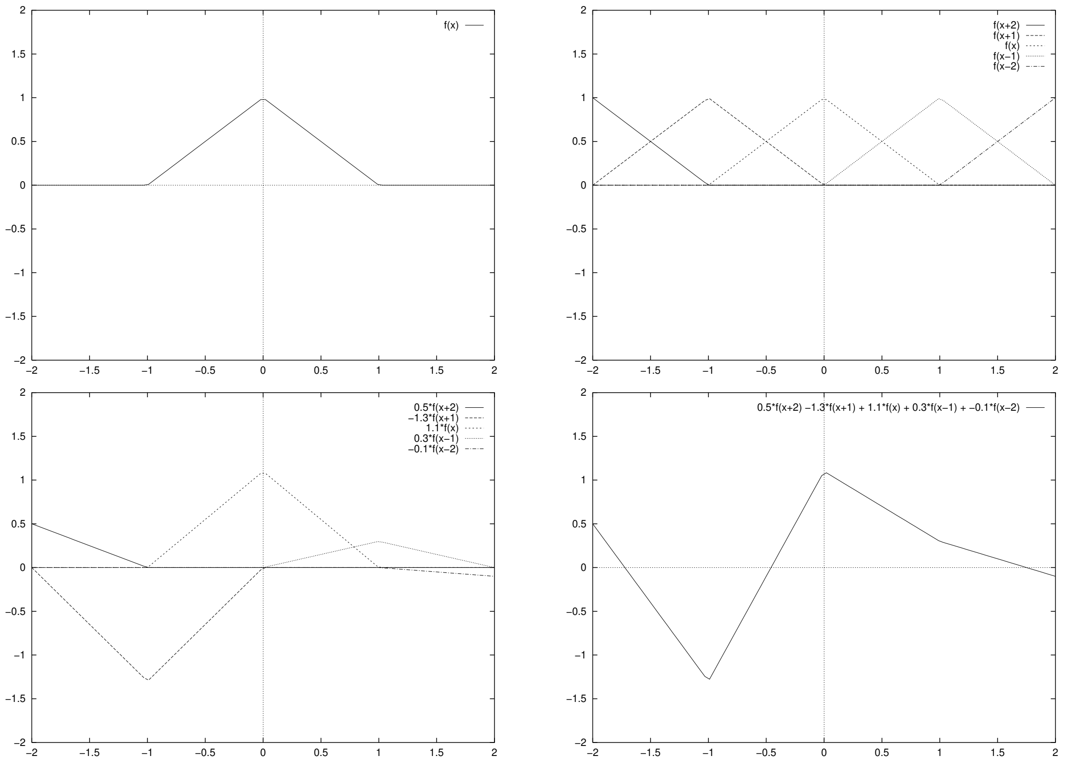


Figure 2.1: The approximation process.

4. We sum these resultant functions.

Figure 2.1 tries to demonstrate this process. The first frame shows our interpolating function, the second shows various translations of it, the third shows these translates scaled by a set of coefficients and the last shows the sum of these. The function which we are trying to approximate might be a polynomial which passes through the points: $(-2, 0.5)$, $(-1, -1.3)$, $(0, 1.1)$, $(1, 0.3)$ and $(2, -0.2)$.

One problem is that our approximation may not be very good. A possible way of improving our approximation is to give ourselves more nodes*. One obvious way of giving ourselves more nodes is to put a new node half way between each of the old nodes. When we've done this we'll probably want to squash up our approximating function too, to stop them overlapping too much.

*Using more nodes doesn't alway work the way we might think. Sometimes it makes things worse! The classic example in this case seems to be due to Runge, and involves approximating $\frac{1}{1+x^2}$ on $[-5, 5]$ with Lagrange polynomials. In this case the approximation oscillates violently for values of $|x|$ which are bigger than about 3.5 — more details can be found in [19].

Essentially this is all there is to multiresolution analysis. Giving ourselves new nodes is a change of *resolution* and squashing our functions is a change of *scale*. Multiresolution analysis studies what happens when we vary this scale.

Now for the formal definition of multiresolution analysis.

Definition 2.1. A multiresolution analysis of $L^2(\mathbb{R})$ is a collection of subsets $\{V_j\}_{j\in\mathbb{Z}}$ of $L^2(\mathbb{R})$ such that:

1. $\exists\, g \in L^2(\mathbb{R})$ so that V_0 consists of all (finite) linear combinations of $\{g(x-k) : k \in \mathbb{Z}\}$,

2. the $g(x-k)$ are an orthonormal † series in V_0,

3. $f(x) \in V_j \iff f(2x) \in V_{j+1}$ ‡,

4. $\bigcup_{j=-\infty}^{+\infty} V_j$ is dense in $L^2(\mathbb{R})$,

5. $\bigcap_{j=-\infty}^{+\infty} V_j = \{0\}$,

6. $V_j \subset V_{j+1}$.

We may think of g as our chosen approximating function. V_0 contains all the functions we can make by adding up translations of g to the integers. When we give ourselves twice as many nodes and squash our generating function we move from V_j to V_{j+1}. Condition 4 means we can approximate any $L^2(\mathbb{R})$ function as closely as we choose. Condition 6 ensures that our approximation is improving all the time: it would be a bit unfortunate if we got to V_{10} and had the exact function we wanted only to find it wasn't in V_{11}.

Some authors leave out the orthogonality requirement (condition 2). They then refer to a multiresolution analysis with this extra property as an orthogonal multiresolution analysis. This would be somewhat clearer, and leave the definition usable on spaces where we do not have the idea of things being orthogonal (for example in $L^p(\mathbb{R})$ for most p). However the majority of authors do include condition 2.

†We say 2 functions f and g are orthogonal if $\int f(x)\overline{g(x)}\,dx = 0$. This parallels two vectors $\vec{x}, \vec{y}$ being orthogonal if $\sum x_i \overline{y_i} = 0$. We say a series is orthonormal if the members are pairwise orthogonal and $\|f\|_2 = 1$ for each f in the sequence.

‡This notation is bad — what we really mean is: $f(.) \in V_j \iff f(2.) \in V_{j+1}$. We will, however, stick with the bad notation, as it seems more natural.

One obvious way to construct an MRA is to pick a g and to produce V_0 using rule 1. Then form V_j using rule 3 and applying rule 4 we, hopefully, get all of $L^2(\mathbb{R})$. We do then have to check that the g which we chose will be orthogonal to its translates, that the intersection of the V_j contains only zero and that $V_j \subset V_{j+1}$.

We can easily extend this definition to $L^2(\mathbb{R}^n)$ for any n. We just replace $L^2(\mathbb{R})$ with $L^2(\mathbb{R}^n)$ throughout the definition and substitute $\{g(x-k) : k \in \mathbb{Z}\}$ with $\{g(x-k) : k \in \mathbb{Z}^n\}$ in condition 1.

2.3 Three Examples

We'll take a look at a few examples which demonstrate a few of the forms a multiresolution analysis can take, and also some possible applications. The first example is the classical example of the Haar multiresolution analysis. This is more or less the canonical multiresolution analysis, and can be kept in mind when dealing with almost any property of a multiresolution analysis.

The second example deals with functions whose Fourier transform is zero outside a given range. It is an interesting MRA because it is easy to describe without defining it in terms of its generating function. Functions whose Fourier transforms are zero outside a given interval are often called *band-limited*.

The last example is generated by one of *Daubechies' generating functions*. This example is of interest because it leads to "compactly supported smooth orthogonal wavelets", which are one of the most touted achievements of the theory of wavelets.

The first two examples are discussed in significant brevity at the beginning of the second chapter of [12], and in [9] in the "Wavelets and Multiresolution Analysis" chapter. Perhaps one of the best references for the third example is the often cited original paper [3], but most introductions [17, 8] to the topic of wavelets will at least mention it.

2.3.1 The Haar MRA

To form the Haar MRA we begin by taking g to be:

$$g(x) = \begin{cases} 1 & x \in [0,1) \\ 0 & \text{otherwise} \end{cases}.$$

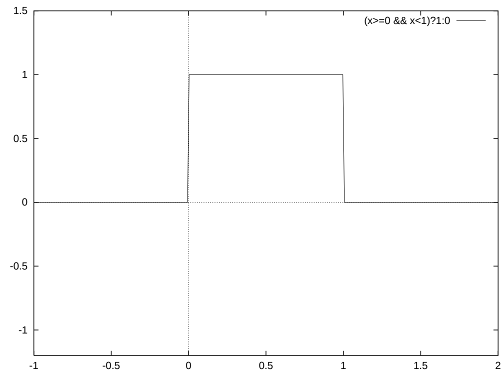


Figure 2.2: $g(x)$ for the Haar MRA.

This is commonly known as the *indicator* or *characteristic* function of the set $[0,1)$, and is usually written $\chi_{[0,1)}(x)$. In this case the translates of g are the characteristic functions of the intervals $[n, n+1)$ for each of the integers n, which are quite clearly orthogonal (as their supports[§] do not even overlap).

We now make V_0 the set of finite linear combinations of these characteristic functions, so V_0 will just contain functions which are piecewise constant on each interval $[n, n+1)$ and have compact support.

We look at what V_1 is going to be. Using $f(x) \in V_0 \iff f(2x) \in V_1$, we find ourselves looking at $g(2x)$, which turns out to be the same as $\chi_{[0,\frac{1}{2})}(x)$. Following from this we find V_1 will contain functions which are constant on $[n/2, n/2 + 1/2)$ where $n \in \mathbb{Z}$. If we repeat this process we find:

$$V_j = \left\{ f : f \text{ has compact support and piecewise constant on } \left[\frac{n}{2^j}, \frac{n+1}{2^j}\right) \forall n \in \mathbb{Z} \right\}.$$

As any function which is constant on $\left[\frac{n}{2^j}, \frac{n+1}{2^j}\right)$ is certainly constant on $\left[\frac{2n}{2^{j+1}}, \frac{2n+1}{2^{j+1}}\right)$, these spaces are nested.

How do we show the union of these V_j is dense in $L^2(\mathbb{R})$? Well, this union contains step functions whose steps change height at $n2^{-j}$ for $n, j \in \mathbb{Z}$. These functions are dense in the set of all step functions, which in turn are dense in $L^2(\mathbb{R})$.

The intersection of all these V_j will contain functions which are constant on intervals of the form $[n2^{-j}, (n+1)2^{-j})$, and which are in $L^2(\mathbb{R})$. In particular by looking at $n = 0$

[§]The support of a function is the closure of the set where is nonzero.

and $n = 1$ we see the functions are constant on $[0, 2^{-j})$ and $[-2^{-j}, 0)$ for any $j \in \mathbb{Z}$. This means that these functions are constant on $[0, \infty)$ and $(-\infty, 0)$. But the only function in $L^2(\mathbb{R})$ which is constant on all such large intervals is the constant zero function.

This multiresolution analysis might be said to consist of functions which are constant on dyadic intervals. A dyadic number is one of the form $n2^{-j}$ with $n, j \in \mathbb{Z}$. Dyadic intervals are intervals with these numbers as end points.

In $L^2(\mathbb{R}^2)$ we could use a similar construction, this time starting with g to be the characteristic function of the unit square $[0,1) \times [0,1)$. In this case we'll get a multiresolution analysis of functions constant on dyadic squares. This naturally generalises to $L^2(\mathbb{R}^n)$.

Figure 2.3 shows an example of how approximations converge to an function on $\mathbb{R}^2$. As this example suggests multiresolution has been applied to various sorts of image analysis. Indeed "mip-mapping" used in image rendering is very similar in structure to this MRA ([22] page 141 and plate 4).

2.3.2 Band-Limited MRA

In this example we start by defining the sets V_j.

$$V_j = \left\{ f \in L^2(\mathbb{R}) : \hat{f} \text{ is supported on } [-2^j\pi, 2^j\pi]. \right\}$$

Here $\hat{f}$ is the Fourier transform of f, which we'll take with the normalisation:

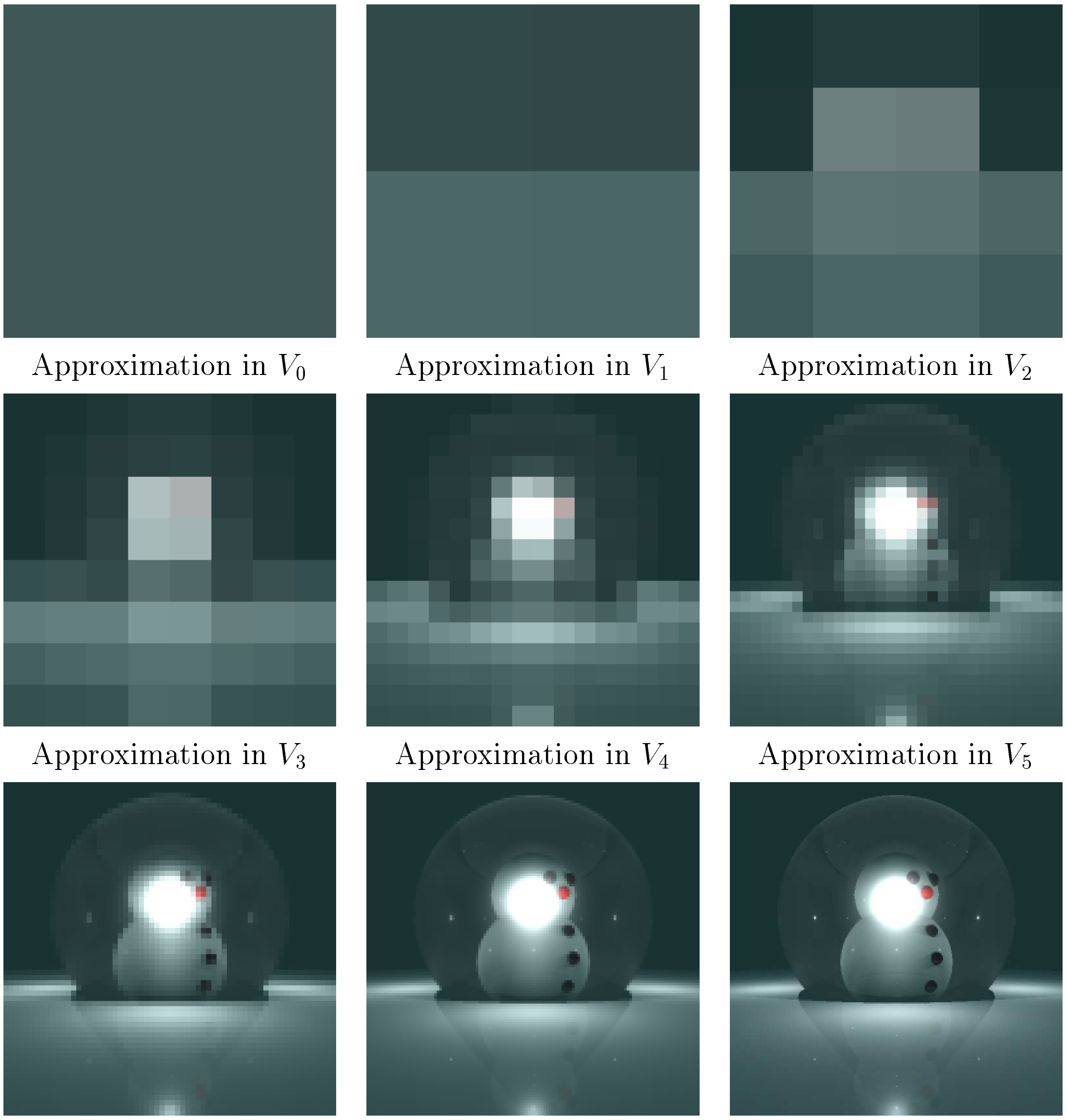
$$\hat{f}(\omega) = \int_{-\infty}^{\infty} e^{-i\omega x} f(x)\, dx.$$

We will also use the notation $\mathcal{F}(f)$ for the Fourier transform of f.

One thing is clear, that as j gets bigger so does V_j, and $V_j \subset V_{j+1}$. Also the fact that the union of these sets is dense in $L^2(\mathbb{R})$ and that the intersection only contains the constant zero function is reasonably clear from the definition of V_j and the fact that $\mathcal{F}$ is both invertible and linear.

To verify that $f(x) \in V_j \iff f(2x) \in V_{j+1}$, we need to look at how $\mathcal{F}(f(x))$ compares to $\mathcal{F}(f(2x))$. We do the calculation, and use the change of variable $x' = 2x$ half way through.

$$\mathcal{F}(f(2x))(\omega) = \int_{-\infty}^{\infty} e^{-i\omega x} f(2x)\, dx = \frac{1}{2} \int_{-\infty}^{\infty} e^{-i\frac{\omega x'}{2}} f(x')\, dx' = \frac{1}{2}\mathcal{F}(f(x))(\omega/2)$$



Approximation in V_0 Approximation in V_1 Approximation in V_2

Approximation in V_3 Approximation in V_4 Approximation in V_5

Approximation in V_6 Approximation in V_7 Approximation in V_8

Figure 2.3: Improving Approximations.

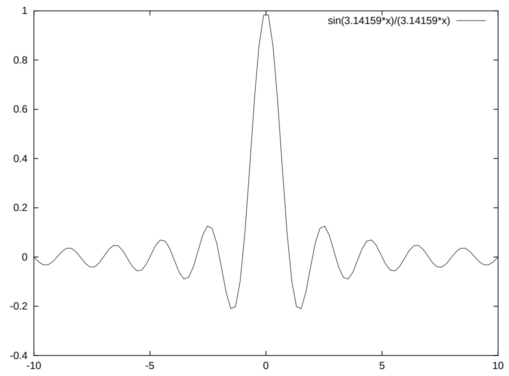


Figure 2.4: $g(x) = \frac{\sin \pi x}{\pi x}$.

From this we can see that if $\mathcal{F}(f(x))$ is supported on $[a,b]$ then $\mathcal{F}(f(2x))$ is supported on $[2a, 2b]$. This is exactly what we need to show $f(x) \in V_j \iff f(2x) \in V_{j+1}$.

Finding a suitable g requires a similar sort of trickery, and is only a little more subtle. This time we need to find out what happens to the Fourier transform as we move from $g(x)$ to $g(x-n)$.

$$\begin{aligned}\mathcal{F}(g(x-n))(\omega) &= \int_{-\infty}^{\infty} e^{-i\omega x} g(x-n)\,dx = \int_{-\infty}^{\infty} e^{-i\omega(x'+n)} g(x')\,dx \\ &= e^{-i\omega n} \int_{-\infty}^{\infty} e^{-i\omega x'} g(x')\,dx = e^{-i\omega n}\mathcal{F}(g(x))(\omega)\end{aligned}$$

This time using the change of variable $x' = x - n$.

We need¶ linear combinations of these $e^{-i\omega n}\hat{g}(\omega)$ to span $\mathcal{F}(V_0)$, which contains all the functions supported on $[-\pi, \pi]$. Now, by considering $\mathcal{F}(V_0)$ as $L^2([-\pi,\pi])$ and using what we already know about Fourier analysis we can get these functions by taking sums of $e^{in\omega}$ with $n \in \mathbb{Z}$. So we just have to multiply by $\chi_{[-\pi,\pi]}$ to kill off anything outside this interval. But this option is open to us! Suppose g has a Fourier transform which is a constant multiple of $\chi_{[-\pi,\pi]}$. Then, from the change of variable formula above, we know the Fourier transform of $g(x-n)$ is just going to be that constant times $e^{-i\omega n}$ on $[-\pi,\pi]$ and zero elsewhere.

It turns out that $g(x) = \frac{\sin \pi x}{\pi x}$ is a g with this property, and what we have been

¶What we really want is translations of g to span V_0, however the invertibility and linearity of $\mathcal{F}$ make this the same as $e^{-i\omega n}\hat{g}(\omega)$ spanning $\mathcal{F}(V_0)$

considering is a special version of *Shannon's Sampling Theorem*, which says the following.

Theorem 2.1. *Let f be a band-limited function, with its Fourier transform supported on $[\Omega/2, \Omega/2]$. If Δ is chosen so that:*

$$\Delta \leq \frac{\pi}{\Omega},$$

then f may be reconstructed exactly from the samples $f_n = f(n\Delta)$ (for $n \in \mathbb{Z}$) by:

$$f(x) = \sum_{n=-\infty}^{\infty} f_n \frac{\sin \pi(\Delta^{-1}x - n)}{\pi(\Delta^{-1}x - n)}.$$

Proof. One proof, not too far from the way we arrived at g ourselves, is found as Theorem 5.1 of [8], however Kaiser's definition of the Fourier transform is slightly different to ours. ■

Let us review the situation. We have been checking this multi-resolution analysis against Definition 2.1. We decided that conditions 6, 5 and 4 were quite believable after examining the definition of V_j. Looking at the relationship between $\mathcal{F}(f(x))$ and $\mathcal{F}(f(2x))$ provided us with what we needed to check condition 3, and Shannon's theorem provided us with a g for condition 1.

This leaves us with only condition 2 to check. It would be reasonable to tackle this problem head on and check:

$$\int g(x)\overline{g(x-m)}\, dx = \int \frac{\sin \pi x}{\pi x} \frac{\sin \pi(x-m)}{\pi(x-m)}\, dx = 0,$$

when $m \neq 0$. However, using another important result from Fourier analysis we can get the result in a simpler manner.

Theorem 2.2. *For $f, g \in L^2(\mathbb{R})$:*

$$\int f(x)\overline{g(x)}\, dx = 2\pi \int \hat{f}(\omega)\overline{\hat{g}(\omega)}\, d\omega.$$

Proof. Almost any book on Fourier analysis will have a proof of this, again [8] provides a discussion of this in section 1.4, but a more theoretical (and terse) discussion can be found in [16] chapter 1, section 2. ■

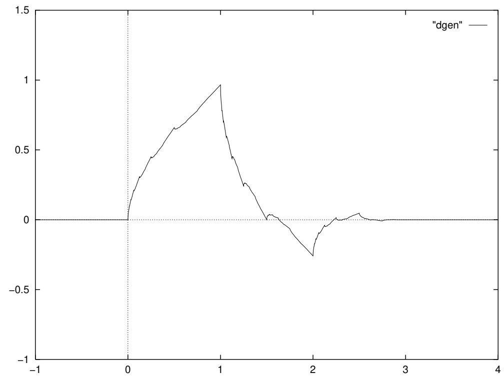


Figure 2.5: $g(x)$ for one of Daubechies' MRAs.

This result is know as Plancherel's theorem (as are several of its variants). It transforms our orthogonality problem into checking:

$$\int_{-\pi}^{\pi} e^{i\omega m} = 0,$$

when $m \neq 0$, which is easy.

In summary, from this example, we have learned a little more about Fourier analysis, and have seen a little of how the structure of a multiresolution analysis interacts with Fourier analysis. It might also make us wonder at all the strange shapes and sizes a multiresolution analysis could take.

2.3.3 Daubechies' generating function

Figure 2.5 shows a rather strange looking function. This function was engineered by Daubechies to have certain properties. It is compactly supported (on $[0,3]$), bounded and generates an MRA which — in some ways — is a near relation of the Haar MRA. Most interestingly both $f(x) = 1$ and $f(x) = x$ can be expressed as a sum of this function's translates — in much the same way as constant functions can be expressed as a sum of the Haar generating function. Consequently this MRA is suitable for the approximation of piecewise linear functions.

Daubechies actually produced a whole family of these generating functions, each smoother than the previous, each supported on a larger interval $[0, 2N-1]$ and able to approximate $1, x, \dots, x^{N-1}$. The key to producing these generating functions was the dilation equation.

These dilation equations in turn lead to a formula for wavelets for the discrete wavelet transform.

2.4 Dilation Equations

While searching for generating functions for a multiresolution analysis it is natural to ask if there are some conditions imposed on the generating function by the structure of the multiresolution analysis. One of the most interesting conditions comes from considering part 6 of the MRA definition — it says $V_0 \subset V_1$, which means:

$$g \in V_0 \subset V_1 = \operatorname{span}\{g(2x-n) : n \in \mathbb{Z}\}.$$

From this we can conclude that:

$$g(x) = \sum_{n\in\mathbb{Z}} c_n g(2x-n).$$

This type of equation, where g is expressed in terms of dilated versions of itself, has been called a *dilation equation*, a *two scale difference equation* or even a *multiscale difference equation*. Solutions to this type of equation are often called *scaling functions* and so the generating function of an MRA is sometimes referred to as the scaling function.

It turns out a lot can be learned about g by examining these coefficients c_n. Conversely by choosing the coefficients carefully we may be able to find a g with certain desirable properties. We will now have a look at how some properties relate to the coefficients.

2.4.1 Integrability

If we suppose that g is integrable (ie. that is in $L^1(\mathbb{R})$) we can integrate over both sides of the dilation equation:

$$\begin{aligned}\int_{\mathbb{R}} g(x)\,dx &= \int_{\mathbb{R}} \sum_{n\in\mathbb{Z}} c_n g(2x-n)\,dx \\ &= \sum_{n\in\mathbb{Z}} c_n \int_{\mathbb{R}} g(2x-n)\,dx \\ &= \sum_{n\in\mathbb{Z}} c_n \frac{1}{2} \int_{\mathbb{R}} g(x)\,dx.\end{aligned}$$

Now, providing $g(x)$ does not have mean zero, we may divide by $\int g(x)\,dx$ to get:

$$2 = \sum_{n\in\mathbb{Z}} c_n.$$

Note that if $\int g(x)\,dx$ diverges we do not get this condition on the c_n.

2.4.2 Orthonormality

All these conditions use similar tricks — this time we begin with the requirement that $g(x)$ and $g(x-n)$ be orthonormal, and then we fill $\sum c_n g(2x-n)$ in for $g(x)$.

$$\begin{aligned}
\delta_{0m} &= \int_{\mathbb{R}} g(x)\overline{g(x-m)}\,dx \\
&= \int_{\mathbb{R}} \left(\sum_{k\in\mathbb{Z}} c_k g(2x-k)\right)\overline{\left(\sum_{l\in\mathbb{Z}} c_l g(2x-l)\right)}\,dx \\
&= \sum_{k\in\mathbb{Z}}\sum_{l\in\mathbb{Z}} c_k\overline{c_l}\int_{\mathbb{R}} g(2x-k)\overline{g(2x-l)}\,dx \\
&= \sum_{k\in\mathbb{Z}}\sum_{l\in\mathbb{Z}} c_k\overline{c_l}\frac{1}{2}\int_{\mathbb{R}} g(x)\overline{g(x+k-l+2m)}\,dx \\
&= \frac{1}{2}\sum_{k\in\mathbb{Z}}\sum_{l\in\mathbb{Z}} c_k\overline{c_l}\delta_{0\,k-l+2m} \\
&= \frac{1}{2}\sum_{k\in\mathbb{Z}} c_k\overline{c_{k+2m}}
\end{aligned}$$

This condition, like the integrability condition, is necessary but may not be sufficient.

2.4.3 Ability to Approximate

Strang, in Appendix 2 of [18], lists the following condition:

$$\sum_{k\in\mathbb{Z}} c_k(-1)^k k^m = 0 \qquad \text{for } m = 0, 1, \ldots, p-1.$$

It has the following following amazing consequences:

1. The polynomials $1, x, \ldots, x^{p-1}$ are linear combinations of $g(x-n)$.

2. Smooth functions can be approximated with error of $O(2^{-pj}\|f^{(p)}\|)$ in V_j.

3. The wavelets we will construct will be orthogonal to $1, x, \ldots, x^{p-1}$. That is:

$$\int x^m w(x)\, dx = 0 \qquad \text{for } m = 0, 1, \ldots, p-1.$$

This integral is sometimes called the m^{th} moment of w.

Proof of these consequences is related to a general theory of approximation by translates developed for the finite element method. The "Strang-Fix" condition relates the goodness of approximation to the degree of the zeros of the Fourier transform of g. This Strang-Fix condition, when applied to the Fourier transform of our dilation equation, produces the condition on the c_n above.

This condition is also mentioned in [5] as being important for the convergence of various schemes for calculating g.

Daubechies' generating function is the unique integrable function which satisfies the Integrability, the Orthonormality and the first 2 approximation conditions (with $m = 0, 1$).

2.5 Discrete Wavelets

Having dealt with dilation equations we are now in a position to build ourselves a discrete wavelet basis for $L^2(\mathbb{R})$. Suppose we are working in some multiresolution analysis, and we are trying to approximate $f \in L^2(\mathbb{R})$. Say that f_j is the best approximation to f in V_j.

As we move from V_j to V_{j+1} our approximation must improve — or at worst stay the same. We can think of this as:

$$f_{j+1} = f_j(x) + d_j(x),$$

where d_j is the extra detail needed to bring our approximation up to the standard in V_{j+1}. Looking at this in a more general context, we might think that:

$$V_{j+1} = V_j \oplus W_j,$$

where the space W_j contains all the details necessary to improve V_j, but is also orthogonal to V_j. Then:

$$V_0 \oplus W_1 \oplus W_2 \oplus W_3 \oplus \cdots$$

will be dense in $L^2(\mathbb{R})$. In fact, in $L^2(\mathbb{R})$:

$$\bigoplus_{j=-\infty}^{\infty} W_j$$

is dense.

Now, the neatness of this scheme becomes apparent when we remember that: $f(x) \in V_j \iff f(2x) \in V_{j+1}$. This allows us to form the same relationship between the W_j as we have between the V_j. If we could also produce a generating function w for W_0 as we produced a g for V_0 then family W_j would have a very neat form indeed.

As luck would have it,

$$w(x) = \sum_{k\in\mathbb{Z}} (-1)^k \overline{c_{1-k}} g(2x-k)$$

is just such a function! We can certainly see that this w is in V_1, as it is the sum of translates of $g(2x)$. We can also check it is not in V_0 by showing that it is orthogonal to all translates of $g(x)$.

$$\begin{aligned}
\int_{\mathbb{R}} g(x-n)w(x)\,dx &= \int_{\mathbb{R}} \left(\sum_{l\in\mathbb{R}} c_l g(2x-l-2n)\right) \overline{\left(\sum_{k\in\mathbb{Z}} (-1)^k \overline{c_{1-k}} g(2x-k)\right)}\,dx \\
&= \sum_{k\in\mathbb{Z}} \sum_{l\in\mathbb{R}} (-1)^k c_{1-k} c_l \int_{\mathbb{R}} g(2x-l-2n)\overline{g(2x-k)}\,dx \\
&= \sum_{k\in\mathbb{Z}} \sum_{l\in\mathbb{R}} (-1)^k c_{1-k} c_l \delta_{k\,l+2n} \qquad \text{as translates are orthonormal} \\
&= \sum_{k\in\mathbb{Z}} (-1)^k c_{1-k} c_{k-2n} \\
&= \sum_{k\in\mathbb{Z}} c_{1-2k} c_{2k-2n} - \sum_{k\in\mathbb{Z}} c_{-2k} c_{2k-2n+1} \quad \text{separating even and odd} \\
&= \sum_{k\in\mathbb{Z}} c_{1-2k} c_{2k-2n} - \sum_{l\in\mathbb{Z}} c_{2l-2n} c_{1-2l} \\
&= 0
\end{aligned}$$

The last step is performed by changing dummy variable to l so that $2k-2n+1 = 1-2l$.

So we know that $w(x)$ is in V_1 and orthogonal to all of V_0, so it seems very likely that w is the required function. For the full proof see [12] Chapter 3 Theorem 1 or the chapter of [9] entitled "Discrete Wavelets and Multiresolution analysis" Theorem 5.1[‖].

This $w(x)$ is our "mother wavelet" which has been produced so that:

$$\left\{w(2^j x - n) : n \in \mathbb{Z}\right\}$$

is an orthogonal basis[**] for W_j. Adjusting these so they are an orthonormal basis, and remembering that $\bigoplus W_j$ is dense in $L^2(\mathbb{R})$ we get the following orthonormal basis:

$$\left\{w_{j\,n}(x) = 2^{\frac{j}{2}} w(2^j x - n) : j, n \in \mathbb{Z}\right\}.$$

We can now say what the discrete wavelet transform is. We have just established that we can write all the functions in $L^2(\mathbb{R})$ in the form:

$$f = \sum_{j,n} a_{j\,n} w_{j\,n},$$

where we can determine the $a_{j\,n}$ using the orthogonality of the $w_{j\,n}$ as follows:

$$a_{j\,n} = \int f(x)\overline{w_{j\,n}(x)}\,dx.$$

This $a_{j\,n}$ is the discrete wavelet transform of f. We can think of 2^{-j} being the frequency and $n2^{-j}$ as being the position, in much the same way as a was the frequency and b was the position in the continuous wavelet transform of Section 1.6.

This discrete transform has quite a lot of advantages. It it less redundant than its continuous relative and it is easier to work with numerically, as all of the practical calculations can be performed with the c_n without ever calculating $w(x)$! Also, the theory is in some way less ad hoc than that for the CWT.

2.6 Back to the examples

Since we have three examples, we should have a look and see what the related dilation equation and wavelet looks like in each case. To find the dilation equation we try to write

[‖]Beware the second author uses a slightly different definition of an MRA!

[**]We have not actually checked this here, but the condition which it produces on the coefficients is the same as that required for the translates of g to be orthonormal in section 2.4.2.

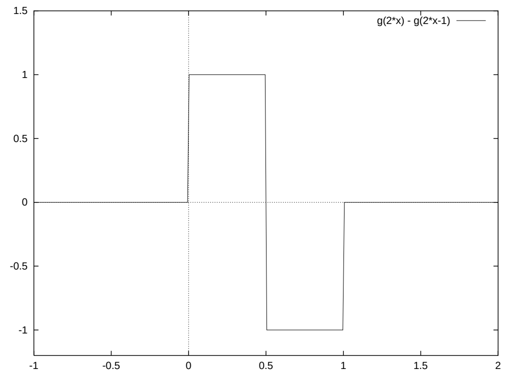


Figure 2.6: $w(x)$ for the Haar MRA.

$g(x)$ in terms of translates of $g(2x)$. To find the wavelet we just use our wavelet formula:

$$w(x) = \sum_{k \in \mathbb{Z}} (-1)^k \overline{c_{1-k}} g(2x - k).$$

2.6.1 Haar Wavelets

In the Haar case we took $g(x) = \chi_{[0,1)}(x)$, and we worked out that $g(2x) = \chi_{[0,\frac{1}{2})}$. In this case it is reasonably clear that:

$$g(x) = \chi_{[0,1)}(x) = \chi_{[0,\frac{1}{2})}(x) + \chi_{[\frac{1}{2},1)}(x) = g(2x) + g(2x - 1).$$

So writing this in the form:

$$g(x) = \sum_{n \in \mathbb{Z}} c_n g(2x - n)$$

We get $c_0 = 1$, $c_1 = 1$ and all the other $c_n = 0$. So using the formula we get:

$$w(x) = \sum_{k \in \mathbb{Z}} (-1)^k \overline{c_{1-k}} g(2x - k) = g(2x) - g(2x - 1).$$

A picture of this Haar wavelet is shown in Figure 2.6.

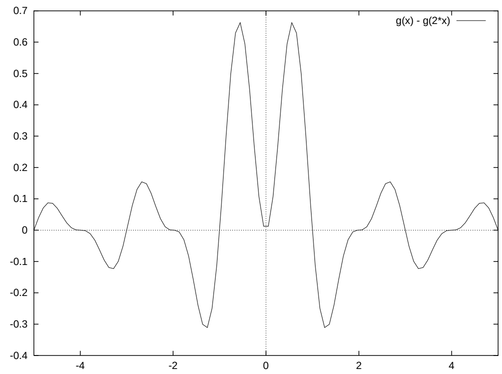


Figure 2.7: $w(x)$ for the Band-Limited MRA.

2.6.2 Shannon Wavelets

In the band-limited MRA we are trying to write $\frac{\sin \pi x}{\pi x}$ in terms of translates of $\frac{\sin 2\pi x}{2\pi x}$. This is going to be a little more complicated than the Haar case. However with the help of Theorem 2.1 with $\Delta = \frac{1}{2}$ we can write:

$$\frac{\sin \pi x}{\pi x} = \sum_n c_n \frac{\sin \pi(2x-n)}{\pi(2x-n)},$$

which is exactly what we require. In this case the c_n are the values of g at $n/2$.

$$c_n = \frac{\sin \pi \frac{n}{2}}{\pi \frac{n}{2}} = \begin{cases} 1 & n = 0 \\ 0 & n \neq 0 \quad \text{is even} \\ \frac{2}{\pi n} & n = 1 \mod 4 \\ -\frac{2}{\pi n} & n = 3 \mod 4 \end{cases}.$$

We could calculate the wavelet using this sequence, but we can form another suitable wavelet more easily. Remember that $\mathcal{F}(g)(\omega)$ was $\chi_{[-\pi,\pi)}(\omega)$, and that $\mathcal{F}(g(2x))(\omega)$ was $\chi_{[-2\pi,2\pi)}(\omega)$. Combining this with Theorem 2.2 we can see that $g(2x) - g(x)$ is orthogonal to $g(x)$, and has all the necessary properties for a wavelet. Figure 2.7 shows a graph of this wavelet.

This wavelet is sometimes called the "Shannon wavelet", because of its close relationship with Shannon's sampling Theorem (Theorem 2.1). It also raises two interesting points. First, sometimes we may have a dilation equation with an infinite number of coefficients.

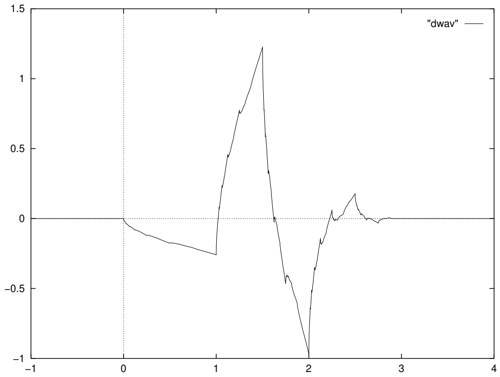


Figure 2.8: $w(x)$ for one of Daubechies MRA's.

Second, the wavelet is not unique — in this case the formula provides a translated version of the wavelet we derived by hand.

2.6.3 Daubechies' Wavelets

As commented earlier: Daubechies generating function was calculated so the coefficients of its dilation equation satisfy all of the following equations.

1. $c_0 + c_1 + c_2 + c_3 = 2$,
2. $c_0\overline{c_2} + c_1\overline{c_3} = 0$,
3. $c_0 - c_1 + c_2 - c_3 = 0$,
4. $0c_0 - 1c_1 + 2c_2 - 3c_4 = 0$.

The values which solve these equations are $c_0 = \frac{1}{4}(1+\sqrt{3})$, $c_1 = \frac{1}{4}(3+\sqrt{3})$, $c_2 = \frac{1}{4}(3-\sqrt{3})$, $c_3 = \frac{1}{4}(1-\sqrt{3})$. Using these coefficients and g we can calculate w (Figure 2.8). In [3] she includes a table of c_n for progressively smoother generating functions (and so wavelets). However the smoother the wavelet the wider its support.

2.7 How to draw these beasties

Until now we have carefully tiptoed around the question of how to draw Daubechies' generating function. The problem is that we are presented with the coefficients for a dilation equation, and left to find the solution ourselves.

Here we will examine an algorithm for drawing solutions to these equations. It makes some rather bold assumptions, but these are justified in the literature [4].

We begin with a set of coefficients $c_0, \dots c_N$, whose sum is 2. We are attempting to sketch a solution g to the dilation equation:

$$g(x) = \sum_{n=0}^{N} c_n g(2x - n).$$

We will assume that $g(x)$ is zero outside $[0, N]$. Now we examine the value of g at each of the integers $0, 1, \dots, N$ using the dilation equation. We arrive at the following relations.

$$\begin{aligned}
g(0) &= c_0 g(0) \\
g(1) &= c_2 g(0) + c_1 g(1) + c_0 g(2) \\
g(2) &= c_4 g(0) + c_3 g(1) + c_2 g(2) + c_1 g(3) + c_0 g(4) \\
g(3) &= c_6 g(0) + c_5 g(1) + c_4 g(2) + c_3 g(3) + c_2 g(1) + \dots \\
&\vdots \\
g(N-1) &= c_N g(N-2) + c_{N-1} g(N-1) + c_{N-2} g(N) \\
g(N) &= c_N g(N)
\end{aligned}$$

This is just an equation of the form:

$$\vec{g} = M\vec{g},$$

where $\vec{g} = (g(0), g(1), \dots, g(N))$ and M is a matrix with the c_n and zeros as entries.

This, however, is just an eigenvector problem, which can be solved either by hand, or by any of a host of computer programs. This provides us with the vector $\vec{g}$ and so the value of g at the integers.

Now that we have found the values at the integers, we can find the values of g at $m/2$ for $m \in \mathbb{Z}$ using the dilation equation:

$$g\left(\frac{m}{2}\right) = \sum_{n=0}^{N} c_n g(m - n),$$

as $m - n$ is an integer. Once we have g at half integers we can get g at quarter integers, eighthes of integers and so on. To draw the graph we just join the dots!

```
#define STEP 0.0002

double c[] = { 0.68301, 1.18301, 0.31699 , −0.18301};

double g(double x)
{
   double tot;
   int i;

   if( x < 0.0 || x > 3.0 ) return 0.0;

   if( fabs(x−0.0) < STEP/2 ) return 0.0;
   if( fabs(x−1.0) < STEP/2 ) return 0.96593;
   if( fabs(x−2.0) < STEP/2 ) return −0.25882;
   if( fabs(x−3.0) < STEP/2 ) return 0.0;

   tot = 0.0;

   for( i = 0 ; i < 4 ; i++ ) tot += c[i] * g(2.0 * x − i);

   return tot;
}
```

Figure 2.9: C code for estimating g

Figure 2.9 contains a piece of C code which shows how simple this scheme is, for the example of Daubechies generating function. The vector $\vec{g}$ was calculated and found to be $(0.0, 0.96593, -0.25882, 0.0)$. This piece of code was used to produce Figure 2.5 and indirectly Figure 2.8.

2.8 Conclusion

We have only scratched the surface of the huge body of material recently produced about the discrete wavelet transform. The same goes for the two auxiliary subjects of multiresolution analysis and dilation equations. A feeling for the volume and variety of work can be got from [6].

Through the examples we have got some idea of the shapes that a multiresolution analysis can come in, and saw that Fourier analysis does seem to have some link to multiresolution analysis. We will see more of this relationship in the following chapters.

One aspect of multiresolution analysis we did not touch on is what Meyer calls the regularity of the MRA. This is a condition on the decay of g and its derivatives. [12] provides great detail on not just this topic, but all of the theory mentioned here.

People have tried to extend wavelet and MRA theory into areas where Fourier analysis has already been used. As long as we stay near $\mathbb{R}^n$ things are very successful: wavelets have provided bases for whole hosts of spaces. Moving to other groups seems to have been more troublesome. In [11] a definition of a multiresolution analysis on a locally compact Abelian group is given, and wavelets are constructed on the Cantor dyadic group — but this sort of work seems to be in the minority.

We have not said much about applications. It has been hinted that discrete wavelets and multiresolution analysis are of use in audio and video signal processing. This is where much of the excitement about wavelets is being generated. Wavelets have been used for both video and audio compression in situations ranging from compressing the FBI's fingerprint collection to transmitting new generation TV signals.

Wavelets have also been put to statistical use for cleaning up data [20], and there is a fast wavelet transform which is even faster than the fast Fourier transform. Some of these many applications can be read about in Appendix 1 and 2 of [18], or in [17], [9] and [6].

Finally, we have said only a little about solving dilation equations. Much of what is known is laid out in [4] and deals with only $L^1(\mathbb{R}^n)$ solutions. In this case the solution is usually unique. An application of dilation equations themselves is proposed in [1]. They examine a signal compression scheme in which the coefficients of a dilation equation, which the signal satisfies, are transmitted in place of the signal. Unfortunately they make incorrect assumptions about the uniqueness of $L^2(\mathbb{R})$ solutions, but the idea still stands.

Chapter 3

A New Construction for the Fourier Transform

3.1 Introduction

Often complicated functions are not defined explicitly, but are rather defined in terms of properties which we know they have. A common example might be the determinant of a matrix. It can be defined either in terms of how to calculate it, or by properties such as what happens when we multiply a row by a constant, or add two rows together. In fact, a determinant can be defined by:

1. $\det : \mathbb{R}^{n\times n} \to \mathbb{R}$
2. $\det(\lambda_1\vec{a}_1, \lambda_2\vec{a}_2, \ldots, \lambda_n\vec{a}_n) = (\lambda_1\lambda_2\ldots\lambda_n)\det(\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_n)$
   where $\lambda_i \in \mathbb{R}$ and $\vec{a}_i \in \mathbb{R}^n$ for $i = 1\ldots n$.
3. $\det(\ldots, \vec{a}_i, \ldots, \vec{a}_j, \ldots) = \det(\ldots, \vec{a}_i + \vec{a}_j, \ldots, \vec{a}_j, \ldots)$
4. $\det(\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n) = 1$ where $\{\vec{e}_j\}_{j=1..n}$ is the usual basis for $\mathbb{R}^n$.

What we will do is give an example of how this may be done for the Fourier transform of functions in $L^2(\mathbb{R})$. The structure on which the determinant is built is the vector space. The framework which we will use for building the Fourier transform is the multiresolution analysis of $L^2(\mathbb{R})$ we introduced in Chapter 2. In this case we're going to take the $\chi_{[0,1)}$ as g and consequently will end up working with the Haar MRA (Section 2.3.1).

The idea for this type of construction came from Chapter 2 of [12], where Meyer often calculates the Fourier transform of a function from its multiresolution expansion. What we do here is show that this calculation may actually be recast as a construction.

There is often debate about factors of $\sqrt{2\pi}$ when defining Fourier transforms, so to begin with the Fourier transform for which we will give a new construction will be the same as that given by:

$$\mathcal{F}(f)(\omega) = \int_{-\infty}^{\infty} e^{-i\omega x} f(x)\, dx.$$

(This is the same as that used in Chapter 2.)

This means that the corresponding inverse transform will have a factor of $1/2\pi$ outside the integral. While we have this formula is in front of us we will do one calculation: the traditional Fourier transform of $\chi_{[0,1)}$

$$\mathcal{F}(\chi_{[0,1)})(\omega) = \int_{-\infty}^{\infty} e^{-i\omega x}\chi_{[0,1)}\, dx = \int_{0}^{1} e^{-i\omega x}\, dx = \frac{1-e^{-i\omega}}{i\omega}$$

3.2 Defining $\mathcal{F}$ on our MRA

First we define $\mathcal{F}$ on $D = \bigcup_{j=-\infty}^{+\infty} V_j$, (with the V_j generated by taking g as $\chi_{[0,1)}$) in the following way:

1. $\mathcal{F} : D \to L^2(\mathbb{R})$ is linear,
2. $\mathcal{F}(f \circ t_n)(\omega) = e^{i\omega n}\mathcal{F}(f)(\omega)$ where $t_n(x) = x + n$,
3. $\mathcal{F}(f \circ d_\lambda)(\omega) = \frac{1}{|\lambda|}\mathcal{F}(f)(\frac{\omega}{\lambda})$ where $d_\lambda(x) = \lambda x$ and $\lambda = 2^n, n \in \mathbb{Z}$,
4. $\mathcal{F}(\chi_{[0,1)})(\omega) = \frac{1-e^{-i\omega}}{i\omega}$.

The linearity, translation and dilation rules are all basic properties of the Fourier transform. The final rule pins down the Fourier transform, hopefully in a way we can extend to all of $L^2(\mathbb{R})$. (We should really check that the translation rule and dilation rule are consistent, that is: $\mathcal{F}(f \circ t_n \circ t_m) = \mathcal{F}(f \circ t_{n+m})$, $\mathcal{F}(f \circ d_\lambda \circ d_\mu) = \mathcal{F}(f \circ d_{\lambda\mu})$ and $\mathcal{F}(f \circ t_m \circ d_\lambda) = \mathcal{F}(f \circ d_\lambda \circ t_{\frac{m}{\lambda}})$.)

Theorem 3.1. *Defining $\mathcal{F}$ using the above rules leads to a well defined function $D \to L^2(\mathbb{R})$.*

Proof. The rules clearly fix $\mathcal{F}$ on all of D, in fact we can provide a formula. If $f \in D = \bigcup V_j$, then we can write f as the finite sum:

$$f(x) = \sum_{k=-N}^{N} a_k \chi_{[0,1)}(2^J x - k),$$

where $f \in V_J$. We can then apply our rules.

$$\begin{aligned}
\mathcal{F}(f)(\omega) &= \sum_{k=-N}^{N} a_k \mathcal{F}(\chi_{[0,1)} \circ t_{-k} \circ d_{2^J})(\omega) \\
&= \sum_{k=-N}^{N} a_k \frac{1}{2^J} \mathcal{F}(\chi_{[0,1)} \circ t_{-k})(\frac{\omega}{2^J}) \\
&= \sum_{k=-N}^{N} a_k \frac{e^{-ik\frac{\omega}{2^J}}}{2^J} \mathcal{F}(\chi_{[0,1)})(\frac{\omega}{2^J}) \\
&= \sum_{k=-N}^{N} a_k \frac{e^{-ik\frac{\omega}{2^J}}}{2^J} \frac{1-e^{-i\frac{\omega}{2^J}}}{i\frac{\omega}{2^J}} \\
&= \frac{1-e^{-i\frac{\omega}{2^J}}}{i\omega} \sum_{k=-N}^{N} a_k e^{-ik\frac{\omega}{2^J}}.
\end{aligned}$$

We are, however, left with some questions.

- Is $\frac{1-e^{-i\omega}}{i\omega}$ in $L^2(\mathbb{R})$? Once we know this we know that $\mathcal{F}(D) \subset L^2(\mathbb{R})$ as $L^2(\mathbb{R})$ is closed under translation, dilation, scaling, addition and multiplication by bounded functions.

- Is $\mathcal{F}$ well defined? Some functions may have more than one expansion as sums of translations and dilations of our basic function $\chi_{[0,1)}(x)$. For example $\chi_{[0,1)} = \chi_{[0,\frac{1}{2})} + \chi_{[\frac{1}{2},1)}$.

- Does $\mathcal{F}$ actually satisfy parts 2 and 3 of its definition on all of D? It is reasonably obvious that $\mathcal{F}$ is linear.

These points are dealt with in the following lemmas. ■

Lemma 3.2.

$$\left\| \frac{1-e^{-i\omega}}{i\omega} \right\|_2^2 = 2\pi$$

Proof. For Theorem 3.1 all we need to show is that the L^2 norm of $\frac{1-e^{-i\omega}}{i\omega}$ is finite, but we can do the calculation exactly using a mixture of brute force and contour integration:

$$\begin{aligned}
\left\| \frac{1-e^{-i\omega}}{i\omega} \right\|_2^2 &= \int_{-\infty}^{\infty} \left| \frac{1-e^{-i\omega}}{i\omega} \right|^2 d\omega \\
&= \int_{-\infty}^{\infty} \frac{(1-e^{-i\omega})(1-e^{i\omega})}{\omega^2} d\omega \\
&= \int_{-\infty}^{\infty} \frac{1-e^{-i\omega}-e^{i\omega}+1}{\omega^2} d\omega \\
&= \int_{-\infty}^{\infty} \frac{2(1-\cos\omega)}{\omega^2} d\omega \\
&= \mathrm{Re}\left(\int_{-\infty}^{\infty} \frac{2(1-e^{i\omega})}{\omega^2} d\omega \right)
\end{aligned}$$

We can now integrate this on a contour which goes to infinity in the upper half plane, and goes around zero on the lower:  The total for this is $(2\pi i)(-2i) = 4\pi$. The contribution of the bit in the upper half plane goes to zero, and the bit around zero contributes 2π as the contour is brought in, so the contribution along the real axis must be 2π. So the total contribution is 2π as required.

■

Lemma 3.3. *$\mathcal{F}$ as defined by the above rule is well defined.*

Proof. First we verify this for the example: $\chi_{[0,1)} = \chi_{[0,\frac{1}{2})} + \chi_{[\frac{1}{2},1)}$.

$$\begin{aligned}
\chi_{[0,1)}(x) &= \chi_{[0,\frac{1}{2})}(x) + \chi_{[\frac{1}{2},1)}(x) \\
&= \chi_{[0,1)}(2x) + \chi_{[1,2)}(2x) \\
&= \chi_{[0,1)}(2x) + \chi_{[0,1)}(2x-1) \\
&= \chi_{[0,1)} \circ d_2(x) + \chi_{[0,1)} \circ t_{-1} \circ d_2(x)
\end{aligned}$$

We already know that: $\mathcal{F}(\chi_{[0,1)})(\omega) = \frac{1-e^{-i\omega}}{i\omega}$, so we now need to work out the Fourier transform of the other two terms:

$$\begin{aligned}\mathcal{F}(\chi_{[0,1)} \circ d_2)(\omega) &= \frac{1}{2}\mathcal{F}(\chi_{[0,1)})\left(\frac{\omega}{2}\right)\\ &= \frac{1}{2}\frac{1-e^{-i\frac{\omega}{2}}}{i\frac{\omega}{2}}\\ &= \frac{1-e^{-i\frac{\omega}{2}}}{i\omega};\end{aligned}$$

$$\begin{aligned}\mathcal{F}(\chi_{[0,1)} \circ t_{-1} \circ d_2)(\omega) &= \frac{1}{2}\mathcal{F}(\chi_{[0,1)} \circ t_{-1})\left(\frac{\omega}{2}\right)\\ &= \frac{1}{2}e^{-i\frac{\omega}{2}}\mathcal{F}(\chi_{[0,1)})\left(\frac{\omega}{2}\right)\\ &= \frac{1}{2}e^{-i\frac{\omega}{2}}\frac{1-e^{-i\frac{\omega}{2}}}{i\frac{\omega}{2}}\\ &= \frac{e^{-i\frac{\omega}{2}}-e^{-i\omega}}{i\omega}.\end{aligned}$$

So we get the required identity as follows:

$$\mathcal{F}(\chi_{[0,\frac{1}{2})} + \chi_{[\frac{1}{2},1)})(\omega) = \frac{1-e^{-i\frac{\omega}{2}}}{i\omega} + \frac{e^{-i\frac{\omega}{2}}-e^{-i\omega}}{i\omega} = \frac{1-e^{-i\omega}}{i\omega} = \mathcal{F}(\chi_{[0,1)})(\omega)$$

Now showing that $\mathcal{F}$ is well defined is much easier, as using this we can break and join adjacent steps. Indeed suppose $f \in D$, then as the V_j are increasing there is a smallest* j for which $f \in V_j$. Then any expansion of f must have the same Fourier transform as the one given by steps of size 2^{-j}, by joining smaller steps. ■

Lemma 3.4. *$\mathcal{F}$ has the following desired properties on all of D:*

$$\mathcal{F}(f \circ t_n)(\omega) = e^{i\omega n}\mathcal{F}(f)(\omega) \quad \textit{for } n \in \mathbb{Z},$$

$$\mathcal{F}(f \circ d_{2^n})(\omega) = \frac{1}{2^n}\mathcal{F}(f)(\frac{\omega}{2^n}) \quad \textit{for } n \in \mathbb{Z}.$$

*If $f \in V_j \forall j$ then $f = 0$, which can be checked easily by hand.

Proof. We simply resort to writing:

$$f(x) = \sum_{k=-N}^{N} a_k \chi_{[0,1)}(2^J x - k),$$

and then using:

$$\mathcal{F}(f)(\omega) = \frac{1 - e^{-i\frac{\omega}{2^J}}}{i\omega} \sum_{k=-N}^{N} a_k e^{-ik\frac{\omega}{2^J}}.$$

As they say "simple algebraic manipulation yields":

$$\begin{aligned}
f \circ t_n(x) &= f(x+n) \\
&= \sum_{k=-N}^{N} a_k \chi_{[0,1)}(2^J(x+n) - k) \\
&= \sum_{l=-N}^{N} a_{l+n2^J} \chi_{[0,1)}(2^J x - l) \\
\Rightarrow \mathcal{F}(f \circ t_n)(\omega) &= \frac{1 - e^{-i\frac{\omega}{2^J}}}{i\omega} \sum_{l=-N}^{N} a_{l+n2^J} e^{-il\frac{\omega}{2^J}} \\
&= \frac{1 - e^{-i\frac{\omega}{2^J}}}{i\omega} \sum_{k=-N}^{N} a_k e^{-i(k-n2^J)\frac{\omega}{2^J}} \\
&= e^{in\omega} \frac{1 - e^{-i\frac{\omega}{2^J}}}{i\omega} \sum_{k=-N}^{N} a_k e^{-ik\frac{\omega}{2^J}} \\
&= e^{in\omega} \mathcal{F}(f)(\omega),
\end{aligned}$$

and:

$$
\begin{aligned}
f \circ d_{2^n}(x) &= f(2^n x) \\
&= \sum_{k=-N}^{N} a_k \chi_{[0,1)}(2^J(2^n x) - k) \\
&= \sum_{k=-N}^{N} a_k \chi_{[0,1)}(2^{J+n} x - k) \\
\Rightarrow \mathcal{F}(f \circ d_{2^n})(\omega) &= \frac{1 - e^{-i\frac{\omega}{2^{J+n}}}}{i\omega} \sum_{k=-N}^{N} a_k e^{-ik\frac{\omega}{2^{J+n}}} \\
&= \frac{1}{2^n} \mathcal{F}(f)(\frac{\omega}{2^n}).
\end{aligned}
$$

■

So now we are at the stage where we have used our rules to define a linear transform $\mathcal{F} : D \to L^2(\mathbb{R})$. We have shown that is it well defined, now it would be nice to extend it to all of $L^2(\mathbb{R})$.

3.3 Extending $\mathcal{F}$ to $L^2(\mathbb{R})$

We know that $D \subset L^2(\mathbb{R})$, and that $\overline{D} = L^2(\mathbb{R})$. We wish to show that $\mathcal{F} : D \to L^2(\mathbb{R})$ can be extended, by taking limits to $\mathcal{F} : L^2(\mathbb{R}) \to L^2(\mathbb{R})$. That is: since for all $f \in L^2(\mathbb{R})$ we may choose a sequence f_n in D so that $f_n \to f$ in $L^2(\mathbb{R})$, we could then define $\mathcal{F}(f)$ as $\lim \mathcal{F}(f_n)$.

If we can show that $\mathcal{F} : D \to L^2(\mathbb{R})$ is continuous then we can use the following lemma.

Lemma 3.5. *Let N and S be Banach spaces. Let M be a dense subset of N. If we have a continuous linear function $\phi : M \to S$, then we can extend ϕ to a function f on N by taking limits, so that f is continuous and linear.*

So extending $\mathcal{F}$ from D to all of $L^2(\mathbb{R})$ comes down to showing that $\mathcal{F} : D \to L^2(\mathbb{R})$ is continuous with the L^2 norm on the domain and range.

Lemma 3.6. *$\mathcal{F} : D \to L^2(\mathbb{R})$ is continuous with the $L^2(\mathbb{R})$ norm on D. In fact:*

$$\|\mathcal{F}(f)\|_2 = \sqrt{2\pi}\|f\|_2$$

Proof. Since $\mathcal{F}$ is linear showing it is bounded, is equivalent to showing it is continuous. Take any $f \in D$, we must show:

$$\|\mathcal{F}(f)\| \leq C\|f\|$$

where C doesn't depend on f. Using our formulation from Theorem 3.1:

$$f(x) = \sum_{k=-N}^{N} a_k \chi_{[0,1)}(2^J x - k).$$

Clearly this gives us $\|f\|_2^2 = \sum_{k=-N}^{N} |a_k|^2/2^J$. Then using the same calculation as in Theorem 3.1 we find $\mathcal{F}(f)$ to be:

$$\frac{1 - e^{-i\frac{\omega}{2^J}}}{i\omega} \sum_{k=-N}^{N} a_k e^{-ik\frac{\omega}{2^J}}.$$

We complete the calculation of $\|\mathcal{F}(f)\|_2^2$ by brute force, contour integration and Lemma 3.2.

$$\begin{aligned}
&\|\mathcal{F}(f)\|_2^2 \\
&= \int_{-\infty}^{\infty} \left| \frac{1 - e^{-i\frac{\omega}{2^J}}}{i\omega} \right|^2 \left(\sum_{k=-N}^{N} a_k e^{-ik\frac{\omega}{2^J}} \right) \left(\sum_{l=-N}^{N} \overline{a_l} e^{il\frac{\omega}{2^J}} \right) d\omega \\
&= \int_{-\infty}^{\infty} \frac{2(1 - \cos\frac{\omega}{2^J})}{\omega^2} \left[\left(\sum_{k=-N}^{N} |a_k|^2 \right) + \left(\sum_{k \neq l} a_k \overline{a_l} e^{-i(k-l)\frac{\omega}{2^J}} \right) \right] d\omega \\
&= \frac{2\pi}{2^J} \sum_{k=-N}^{N} |a_k|^2 + \sum_{k \neq l} a_k \overline{a_l} \int_{-\infty}^{\infty} \frac{2(1 - \cos\frac{\omega}{2^J})}{\omega^2} e^{-i(k-l)\frac{\omega}{2^J}} \, d\omega.
\end{aligned}$$

We want the integral on the right to be zero. By a change of variable, we reduce the problem to showing that:

$$\int_{-\infty}^{\infty} \frac{2(1 - \cos\omega)}{\omega^2} e^{ir\omega} \, d\omega = 0,$$

when $r \in \mathbb{Z}, r \neq 0$. The imaginary part has to be zero, as $\frac{2(1-\cos\omega)}{\omega^2}$ is even and $\sin r\omega$ is odd. So now we're down to:

$$\begin{aligned}
&\int_{-\infty}^{\infty} \frac{2(1-\cos\omega)}{\omega^2}\cos r\omega \, d\omega \\
&= \int_{-\infty}^{\infty} \frac{2\cos r\omega - 2\cos\omega\cos r\omega}{\omega^2}\, d\omega \\
&= \int_{-\infty}^{\infty} \frac{2\cos r\omega - \cos((r+1)\omega) - \cos((r-1)\omega)}{\omega^2}\, d\omega \\
&= \operatorname{Re}\left(\int_{-\infty}^{\infty} \frac{2e^{ir\omega} - e^{i(r+1)\omega} - e^{i(r-1)\omega}}{\omega^2}\, d\omega\right).
\end{aligned}$$

Now, it's easy to check that $2e^{ir\omega} - e^{i(r+1)\omega} - e^{i(r-1)\omega}$ has a double zero at $\omega = 0$, and so the integrand is analytic at zero. If $r \leq -1$ or $r \geq 1$ then we can integrate around a loop in the lower or upper half plane respectively, so that the contribution for this part of the loop goes to zero. But as the function is analytic everywhere, the integral around the whole loop must be zero, and contribution from integrating along the real line must be zero.

This means that as $\|f\|_2^2 = \sum_{k=-N}^{N} |a_k|^2/2^J$ we simply get $\|\mathcal{F}(f)\|_2^2 = 2\pi\|f\|_2^2$. So we see $\mathcal{F}$ is bounded with norm $\sqrt{2\pi}$. ■

Now we can extend $\mathcal{F}$ as was desired, by combining the previous results.

Theorem 3.7. *$\mathcal{F} : D \to L^2(\mathbb{R})$ can be extended continuously to $\mathcal{F} : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ in a unique way. Also this extension satisfies $\|\mathcal{F}(f)\|_2 = \sqrt{2\pi}\|f\|_2$ for $f \in L^2(\mathbb{R})$.*

Proof. Use Lemma 3.6 to show $\mathcal{F}$ is bounded and then use Lemma 3.5 extend $\mathcal{F}$. The fact that the norm is just scaled comes from Lemma 3.6, and the fact D is dense in $L^2(\mathbb{R})$. ■

We only needed $\|\mathcal{F}(f)\|_2^2 \leq 2\pi\|f\|_2^2$ and so this stronger result is really just Plancherel's Theorem.

Corollary 3.8 (Plancherel's Theorem). *For $f, g \in L^2(\mathbb{R})$:*

$$(f, g) = 2\pi(\mathcal{F}(f), \mathcal{F}(g))$$

where $(\cdot,\cdot)$ is the usual inner product on $L^2(\mathbb{R})$.

Proof. The fact that the norm is just scaled means we can apply the polarisation identity, in the usual way, to show that the inner product is preserved in this Hilbert space. ■

3.4 Back to the traditional

We are now in a position to get the traditional Fourier transform, from our new definition. This is a rather useful thing, as it allows us to check that no gremlins have crept in and made our new transform different to the old.

Theorem 3.9. *If $f \in L^2(\mathbb{R})$ and*

$$\int_{-\infty}^{\infty} f(x)e^{-i\omega x}\,dx$$

converges for almost every value of ω then it agrees with $\mathcal{F}(f)$ almost everywhere.

Proof. We'll do this in 3 stages, first for functions in D, then for compactly supported functions in $L^2(\mathbb{R})$, and finally for the general case in the statement.

For $f \in D$: Suppose $f \in V_J$ is supported on $[-R, R]$. Then we can write $f(x)$ as:

$$f(x) = \sum_{k=-R2^J}^{R2^J-1} a_k \chi_{[0,1)}(2^J x - k)$$

Now $f \in V_j, \forall j \geq J$, so by noting that the coefficient of $\chi_{[0,1)}(2^j x - k)$ will just be $f(\frac{k}{2^j})$, we may also write f as:

$$f(x) = \sum_{k=-R2^j}^{R2^j-1} f(\frac{k}{2^j})\chi_{[0,1)}(2^j x - k)$$

We take the Fourier transform of this, which we can write out using the calculation in Theorem 3.1.

$$\begin{aligned}
\mathcal{F}(f)(\omega) &= \frac{1-e^{-i\frac{\omega}{2^j}}}{i\omega} \sum_{k=-R2^j}^{R2^j-1} f(\frac{k}{2^j})e^{-ik\frac{\omega}{2^j}} \\
&= \left(\frac{1-e^{-i\frac{\omega}{2^j}}}{i\frac{\omega}{2^j}}\right)\left(\frac{1}{2^j}\sum_{k=-R2^j}^{R2^j-1} f(\frac{k}{2^j})e^{-ik\frac{\omega}{2^j}}\right)
\end{aligned}$$

Fixing ω we look at what happens as $j \to \infty$. The first term goes to 1. Remembering that f is a step function, and so Riemann integrable, the second term becomes a

Riemann sum, leaving us with:

$$\mathcal{F}(f)(\omega) = \int_{-R}^{R} f(x)e^{-i\omega x}\, dx.$$

For $g \in L^2(\mathbb{R})$ with compact support: From here it is easy to extend this to $g \in L^2(\mathbb{R})$ with g supported on $[-R, R]$. First note that if we define

$$\mathcal{F}_R(f) = \int_{-R}^{R} f(x)e^{-i\omega x}\, dx,$$

then $\mathcal{F}_R(\cdot)(\omega)$ is continuous on $L^2(\mathbb{R})$, because:

$$\begin{aligned} \mathcal{F}_R(f)(\omega) &= (f(x), e^{ixw}\chi_{[-R,R)}) \\ \Rightarrow |\mathcal{F}_R(f)(\omega)| &\leq \|f\|_2 \left\| e^{ixw}\chi_{[-R,R)} \right\|_2 \\ &= \|f\|_2 \sqrt{2R}. \end{aligned}$$

This means if a sequence of functions g_n converges in $L^2(\mathbb{R})$ then the sequence of functions $\mathcal{F}_R(g_n)$ has a pointwise limit.

We can choose $g_n \in D$ supported on $[-R, R]$ so that $g_n \to g$ in $L^2(\mathbb{R})$ as $n \to \infty$ and we now know that $\mathcal{F}_R(g_n)$ has a pointwise limit. Looking in $L^2(\mathbb{R})$:

$$\lim_{n\to\infty} \mathcal{F}_R(g_n) = \lim_{n\to\infty} \mathcal{F}(g_n) = \mathcal{F}(g),$$

where all the limits are in $L^2(\mathbb{R})$. Lemma 3.10 tells us that the pointwise limit must agree with $\mathcal{F}(g)$ almost everywhere. That is:

$$\int_{-R}^{R} g(x)e^{-i\omega x}\, dx = \mathcal{F}(g)(\omega) \quad \text{a.e.} \quad \omega$$

So this Fourier transform agrees with the traditional one on functions in $L^2(\mathbb{R})$ with compact support. As g is supported on $[-R, R]$ we can just write this as:

$$\mathcal{F}(g)(\omega) = \int_{-\infty}^{\infty} g(x)e^{-i\omega x}\, dx \quad \text{a.e.} \quad \omega \in \mathbb{R}.$$

What if the formula looks OK? If f is any function in $L^2(\mathbb{R})$ then the simple sequence $f_n = f\chi_{[-n,n)}$ converges to f. What we would like to do is apply the formula to these to get an indication of the Fourier transform of f. This leads us in the direction of the improper integral we want.

$$\begin{aligned}\mathcal{F}(f_n)(\omega) &= \int_{-\infty}^{\infty} f_n(x)e^{-i\omega x}\,dx \quad \text{a.e.} \quad \omega \\ &= \int_{-\infty}^{\infty} \chi_{[-n,n)}(x)f(x)e^{-i\omega x}\,dx \\ &= \int_{-n}^{n} f(x)e^{-i\omega x}\,dx\end{aligned}$$

We want to take a limit of this as $n \to \infty$. If we keep away from the points where this isn't true for some n (which is a countable union of sets of zero measure, and so of zero measure), then we get:

$$\lim_{n\to\infty} \mathcal{F}(f_n)(\omega) = \int_{-\infty}^{\infty} f(x)e^{-i\omega x}\,dx \quad \text{a.e.} \quad \omega$$

The integral on the right is an improper integral, but that isn't the problem. This is a *pointwise* limit, and we need an $L^2(\mathbb{R})$ limit.

All is not lost however. We again use Lemma 3.10 as we have an $L^2(\mathbb{R})$ limit (the Fourier transform) and a pointwise limit of the same sequence of functions and conclude these are be the same almost everywhere. That is: we know $\mathcal{F}(f_n) \to \mathcal{F}(f)$ in $L^2(\mathbb{R})$ and this means that if our improper integral converges to some function almost everywhere, then it is the Fourier transform of f.

■

Lemma 3.10. *Suppose $f_n \to f$ in $L^2(\mathbb{R})$ and $f_n(x) \to g(x)$ pointwise for $x \in \mathbb{R} \setminus N$ where N has measure zero. Then $f(x) = g(x)$ a.e. $x \in \mathbb{R}$.*

Proof. First note that $\forall \delta > 0, h \in L^2(\mathbb{R})$:

$$\|h\|_2^2 \geq \int_{\{x:|h(x)|>\delta\}} \delta^2\,dx = \delta^2 \left|\{x : |h(x)| > \delta\}\right|.$$

This means that:

$$\left|\{x : |f_n(x) - f(x)| > \delta\}\right| \leq \frac{\|f_n - f\|_2^2}{\delta^2}.$$

As $f_n \to f$ in $L^2(\mathbb{R})$ we may choose an increasing sequence n_j such that:

$$\left|\left\{x : |f_n(x) - f(x)| > \frac{1}{2^j}\right\}\right| \leq 2^{2j}\|f_n - f\|_2^2 \leq \frac{1}{2^j},$$

when $n > n_j$, by making $\|f_n - f\|_2^2 \leq 2^{-3j}$. Set:

$$M_j = \bigcup_{k>j} \left\{x : |f_{n_k}(x) - f(x)| > \frac{1}{2^k}\right\},$$

then $|M_j| \leq \frac{1}{2^j}$. Now off this set M_j we know that:

$$|f_{n_k}(x) - f(x)| \leq \frac{1}{2^k}$$

if $k > j$. Thus $f_{n_k}(x) \to f(x)$ pointwise off M_j, but remember that $f_{n_k}(x)$ must also go to $g(x)$ off N, so we conclude $f(x) = g(x)$ off $M_j \cup N$.

Finally let $M = \{x : g(x) \neq f(x)\}$, then $M \subset M_j \cup N$ for all $j \in \mathbb{N}$. Thus $|M| \leq |M_j| + |N| \leq \frac{1}{2^j} + 0$ for all $j \in \mathbb{N}$, so $|M| = 0$ as required. ■

Now we have the advantage that if we want to use a traditional proof we can, as we have the integral formula for just about all the cases we could hope for!

3.5 Can we do better than that?

We defined the Fourier transform from the following 4 rules.

1. $\mathcal{F} : D \to L^2(\mathbb{R})$ is linear,
2. $\mathcal{F}(f \circ t_n)(\omega) = e^{i\omega n}\mathcal{F}(f)(\omega)$ where $t_n(x) = x + n$,
3. $\mathcal{F}(f \circ d_\lambda)(\omega) = \frac{1}{|\lambda|}\mathcal{F}(f)(\frac{\omega}{\lambda})$ where $d_\lambda(x) = \lambda x$ and $\lambda = 2^n, n \in \mathbb{Z}$,
4. $\mathcal{F}(\chi_{[0,1)})(\omega) = \frac{1-e^{-i\omega}}{i\omega}$.

Surprisingly the fourth rule, which seems to kick start this definition of $\mathcal{F}$, may be made weaker. This is because we had to check that $\mathcal{F}$ was well defined in Lemma 3.3, which may be turned around and made into deductions about what $\mathcal{F}(\chi_{[0,1)})$ could possibly be.

Theorem 3.11. *Using the first three rules above and the following:*

- *$\mathcal{F}(\chi_{[0,1)})(\omega)$ is continuous at 0 and $\mathcal{F}(\chi_{[0,1)})(0) = 1$*

we may derive the fourth rule.

Proof. When we checked that $\mathcal{F}$ was well defined we made sure this rule was consistent with $\chi_{[0,1)} = \chi_{[0,\frac{1}{2})} + \chi_{[\frac{1}{2},1)}$. For this to be true we required:

$$\mathcal{F}(\chi_{[0,1)})(\omega) = \frac{1}{2}\mathcal{F}(\chi_{[0,1)})\left(\frac{\omega}{2}\right) + \frac{1}{2}e^{-i\frac{\omega}{2}}\mathcal{F}(\chi_{[0,1)})\left(\frac{\omega}{2}\right)$$

So letting $f = \mathcal{F}(\chi_{[0,1)})$:

$$\begin{aligned} f(\omega) &= \frac{1}{2}\left(1 + e^{-i\frac{\omega}{2}}\right) f(\frac{\omega}{2}) \\ &= \frac{1}{2^2}\left(1 + e^{-i\frac{\omega}{2}}\right)\left(1 + e^{-i\frac{\omega}{2^2}}\right) f(\frac{\omega}{2^2}) \\ &= \left(\frac{1}{2^n}\prod_{k=1}^{n}\left(1 + e^{-i\frac{\omega}{2^k}}\right)\right) f(\frac{\omega}{2^n}) \\ &= \left(\frac{1}{2^n}\sum_{r=1}^{2^n} e^{-i\frac{\omega}{2^n}r}\right) f(\frac{\omega}{2^n}) \\ &= \frac{1}{2^n}\frac{e^{-i\omega} - 1}{e^{-i\frac{\omega}{2^n}} - 1} f(\frac{\omega}{2^n}). \end{aligned}$$

Our new rule supposes $f(\omega)$ is continuous at zero, then (for fixed ω) we take the limit of this expression as $n \to \infty$.

$$f(\omega) = \frac{1 - e^{-i\omega}}{i\omega} f(0)$$

We also assumed that $f(0) = 1$, so:

$$\mathcal{F}(\chi_{[0,1)})(\omega) = f(\omega) = \frac{1 - e^{-i\omega}}{i\omega}$$

So we have derived the old fourth rule as required.

■

This new rule could be motivated by the fact that $\chi_{[0,1)}$ is an $L^1(\mathbb{R})$ function, and if we are aiming for the traditional Fourier transform we want L^1 functions to go to bounded continuous functions.

We can't do any better that this without changing the first 3 rules. Clearly we need something to determine $\mathcal{F}$ up to constant multiples; something like $\mathcal{F}(\chi_{[0,1)})(0) = 1$ or $\|\mathcal{F}(\chi_{[0,1)})\|_2^2 = 2\pi$. If we also drop the continuity at 0 condition we leave open options such as:

$$\operatorname{sign}(\omega)\frac{1-e^{-i\omega}}{i\omega} \qquad \text{where} \quad \operatorname{sign}(x) = \begin{cases} +1 & x > 0 \\ 0 & x = 0 \\ -1 & x < 0 \end{cases}$$

for $\mathcal{F}(\chi_{[0,1)})$. It is easy to see that this will be consistent with the first 3 properties, because they do not relate $\mathcal{F}(\chi_{[0,1)})$ at negative ω with values at positive ω.

If we change the third rule to allow all $\lambda \neq 0$, then we may stand a better chance, as it will relate $\mathcal{F}(\chi_{[0,1)})(\omega)$ for negative and irrational ω to $\mathcal{F}(\chi_{[0,1)})$ on positive dyadic rational ω. This is pursued further in Chapter 4.

3.6 Extending properties of $\mathcal{F}$

In Section 3.2 we defined $\mathcal{F}$ to have certain properties on D which we checked in Lemma 3.4. We said what effect translation and dilation had on $\mathcal{F}$ when we dilated by two or translated by an integer. We will extend these properties in two ways. First we will allow translation and dilation by any real number (except dilation by 0). Second we will show these properties hold on all on $L^2(\mathbb{R})$ and not just on D. We will also establish another related property of the Fourier transform.

Using Theorem 3.9 we easily can extend our translation and dilation rules to all translations t_α with $\alpha \in \mathbb{R}$ and dilations d_λ with $\lambda \in \mathbb{R}\setminus\{0\}$. This just involves a simple change of variable in the integral formula.

In a similar manner we can also examine the effect of translation on the Fourier transform of a function. If $f \in D$, then f has compact support, and so we know:

$$\mathcal{F}(f)(\omega) = \int_{-\infty}^{\infty} f(x)e^{-i\omega x}\,dx$$

This lets us deduce that:

$$\begin{aligned}\mathcal{F}(f)(\omega+\alpha) &= \int_{-\infty}^{\infty} f(x)e^{-i(\omega+\alpha)x}\,dx \quad \text{a.e.} \quad \omega \\ &= \int_{-\infty}^{\infty} f(x)e^{-i\alpha x}e^{-i\omega x}\,dx \quad \text{a.e.} \quad \omega \\ &= \mathcal{F}(f(x)e^{-i\alpha x})(\omega) \quad \text{a.e.} \quad \omega\end{aligned}$$

So $\mathcal{F}(f)\circ t_\alpha = \mathcal{F}(f(x)e^{-i\alpha x})$ when $f \in D$.

To express all this clearly, we'll define 3 operators. These are translation $\mathcal{T}$, dilation $\mathcal{D}$ and rotation $\mathcal{R}$. Our reason for defining these operators is that the properties of $\mathcal{F}$ can be expressed in terms of how $\mathcal{F}$ commutes with them. This notation is clearer than the "$f \circ t_\alpha$" notation above, and as all these operators are continuous it makes it easy to extend a property of $\mathcal{F}$ on D to a property of $\mathcal{F}$ on $L^2(\mathbb{R})$.

- For all $\alpha \in \mathbb{R}$ we define $\mathcal{T}_\alpha : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ by $(\mathcal{T}_\alpha f)(x) = f(x+\alpha)$. This is a continuous linear operator with norm 1 on $L^2(\mathbb{R})$.
- For all $\lambda \in \mathbb{R}\setminus\{0\}$ we define $\mathcal{D}_\lambda : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ by $(\mathcal{D}_\lambda f)(x) = f(\lambda x)$. This is a continuous linear operator with norm $|\lambda|^{-\frac{1}{2}}$.
- For all $\alpha \in \mathbb{R}$ we define $\mathcal{R}_\alpha : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ by $(\mathcal{R}_\alpha f)(x) = e^{i\alpha x}f(x)$. This is also continuous linear operator with norm 1.

Let us now rephrase $\mathcal{F}$'s properties in terms of these operators. For all $f \in D$, $\alpha \in \mathbb{R}$ and $\lambda \in \mathbb{R}\setminus\{0\}$:

1. $\mathcal{F}\mathcal{T}_\alpha f = \mathcal{R}_\alpha \mathcal{F} f$,
2. $\mathcal{F}\mathcal{D}_\lambda f = \frac{1}{|\lambda|}\mathcal{D}_{\frac{1}{\lambda}}\mathcal{F} f$,
3. $\mathcal{F}\mathcal{R}_\alpha f = \mathcal{T}_{-\alpha}\mathcal{F} f$.

So all the properties can be given in terms of relations of the form $\mathcal{F}\mathcal{A} = \mathcal{B}\mathcal{F}$. This gives us a simple way to extend all these properties to $L^2(\mathbb{R})$ in one fell swoop. Suppose f is in $L^2(\mathbb{R})$, and $\mathcal{A}, \mathcal{B}$ are bounded linear operators on $L^2(\mathbb{R})$, such that $\mathcal{F}\mathcal{A} = \mathcal{B}\mathcal{F}$ on D. We can just take $\{f_n\} \subset D$ so $f_n \to f$ as $f_n \to \infty$, giving:

$$\mathcal{F}\mathcal{A}f = \mathcal{F}\mathcal{A}\lim_{n\to\infty} f_n = \lim_{n\to\infty}\mathcal{F}\mathcal{A}f_n = \lim_{n\to\infty}\mathcal{B}\mathcal{F}f_n = \mathcal{B}\lim_{n\to\infty}\mathcal{F}f_n = \mathcal{B}\mathcal{F}f$$

3.7 The inverse Fourier transform

In exactly the same manner as we did for $\mathcal{F}$ we may define another transform $\mathcal{G}$. We want this to be an inverse for the Fourier transform so this time we define it with the following rules:

1. $\mathcal{G} : D \to L^2(\mathbb{R})$ is linear,
2. $\mathcal{G}(f \circ t_n)(\omega) = e^{-i\omega n}\mathcal{G}(f)(\omega)$ where $t_n(x) = x + n$,
3. $\mathcal{G}(f \circ d_\lambda)(\omega) = \frac{1}{|\lambda|}\mathcal{G}(f)(\frac{\omega}{\lambda})$ where $d_\lambda(x) = \lambda x$ and $\lambda = 2^n, n \in \mathbb{Z}$,
4. $\mathcal{G}(\chi_{[0,1)})(\omega)$ is continuous at 0 and $\mathcal{G}(\chi_{[0,1)})(0) = \frac{1}{2\pi}$.

These are the translation and dilation properties we would expect the inverse of the Fourier transform to have (compare to those on page 37). It is easily guessed that $\mathcal{G}$ is very similar to $\mathcal{F}$, but with norm $\frac{1}{\sqrt{2\pi}}$ on $L^2(\mathbb{R})$ and slightly different properties. We would hope that this turns out to be the inverse of the Fourier transform.

As we did for $\mathcal{F}$, we write the properties of this new extended transform in terms of the operators $\mathcal{T}_\alpha$, $\mathcal{D}_\lambda$ and $\mathcal{R}_\alpha$. For all $f \in D$, $\alpha \in \mathbb{R}$ and $\lambda \in \mathbb{R} \setminus \{0\}$:

1. $\mathcal{G}\mathcal{T}_\alpha f = \mathcal{R}_{-\alpha}\mathcal{G}f$,
2. $\mathcal{G}\mathcal{D}_\lambda f = \frac{1}{|\lambda|}\mathcal{D}_{\frac{1}{\lambda}}\mathcal{G}f$,
3. $\mathcal{G}\mathcal{R}_\alpha f = \mathcal{T}_\alpha\mathcal{G}f$.

Compare these to the properties of $\mathcal{F}$ above — the dilation rule is the same, and the minus sign has moved from the rotation rule to the translation rule.

Theorem 3.12. *The Fourier transform $\mathcal{F} : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ is invertible, with inverse $\mathcal{G}$ defined as above.*

Proof. We do already know that $\mathcal{F}$ is injective from Lemma 3.6. However, rather than show it is surjective by hand, it is easier to show that $\mathcal{G}$ is the inverse of $\mathcal{F}$ on $L^2(\mathbb{R})$.

We examine $\mathcal{I} = \mathcal{F}\mathcal{G}$. From the fact that $\mathcal{F}$ and $\mathcal{G}$ are continuous we see that $\mathcal{I}$ is a continuous linear operator in $L^2(\mathbb{R})$ with norm 1 (the fact $\mathcal{F}$ scales the norm by $\sqrt{2\pi}$, and $\mathcal{G}$ scales it by $1/\sqrt{2\pi}$, means that $\mathcal{I}$ preserves the norm). Examining the effect of translation and dilation:

$$\mathcal{I}\mathcal{T}_n f = \mathcal{F}\mathcal{G}\mathcal{T}_n f = \mathcal{F}\mathcal{R}_{-n}\mathcal{G}f = \mathcal{T}_n\mathcal{F}\mathcal{G}f = \mathcal{T}_n\mathcal{I}f$$

$$\mathcal{I}\mathcal{D}_\lambda f = \mathcal{F}\mathcal{G}\mathcal{D}_\lambda f = \frac{1}{|\lambda|}\mathcal{F}\mathcal{D}_{\frac{1}{\lambda}}\mathcal{G}f = \frac{|\lambda|}{|\lambda|}\mathcal{D}_\lambda\mathcal{F}\mathcal{G}f = \mathcal{D}_\lambda\mathcal{I}f$$

So both translation and dilation commute with $\mathcal{I}$. This, combined with the fact that $\mathcal{I}$ preserves the norm, is a pretty strong indication that it may be the identity. If we can show that our generating function $\chi_{[0,1)}$ is sent to itself by $\mathcal{I}$ we know it is the identity.

By using the equivalent of Theorem 3.9 for $\mathcal{G}$ we try to calculate $\mathcal{G}(\frac{1-e^{-i\omega}}{i\omega})$. As an $L^2(\mathbb{R})$ function $\frac{1-e^{-i\omega}}{i\omega}$ must have an image under $\mathcal{G}$, and if the integral below converges to an $L^2(\mathbb{R})$ function, then this must be the image of $\frac{1-e^{-i\omega}}{i\omega}$.

$$\begin{aligned}
&\int_{-\infty}^{\infty} \frac{1-e^{-i\omega}}{i\omega} e^{i\omega x}\, d\omega \\
&= \int_{-\infty}^{\infty} \frac{e^{i\omega x} - e^{i\omega(x-1)}}{i\omega}\, d\omega \\
&= \int_{-\infty}^{\infty} \frac{\sin\omega x - \sin\omega(x-1)}{\omega}\, d\omega \\
&= \int_{-\infty}^{\infty} \frac{\sin\omega x}{\omega}\, d\omega - \int_{-\infty}^{\infty} \frac{\sin\omega(x+1)}{\omega}\, d\omega \\
&= \operatorname{sign}(x)\int_{-\infty}^{\infty} \frac{\sin\omega}{\omega}\, d\omega - \operatorname{sign}(x-1)\int_{-\infty}^{\infty} \frac{\sin\omega}{\omega}\, d\omega \\
&= (\operatorname{sign}(x) - \operatorname{sign}(x-1))\,\pi \\
&= 2\pi\chi_{[0,1)} \quad \text{a.e.} \quad \omega
\end{aligned}$$

When we put in the $\frac{1}{2\pi}$ factor for $\mathcal{G}$ we get exactly what we wanted. Similarly we can check $\mathcal{I}' = \mathcal{G}\mathcal{F}$ is also the identity and so we conclude that $\mathcal{G} = \mathcal{F}^{-1}$ and so $\mathcal{F}$ is invertible. ■

From now on we will use $\mathcal{F}^{-1}$ for $\mathcal{G}$ when talking about the inverse of the Fourier transform on $L^2(\mathbb{R})$.

3.8 Higher dimensions

This definition works quite well in n dimensions too. There are really only two differences. First the dilation equation becomes more complicated — instead of 2 terms we have 2^n

terms. The second difference is that for each dimension we get a factor of

$$\frac{1 - e^{-i\omega}}{i\omega}$$

in the Fourier transform of the function generating the MRA. This is actually quite fortunate, because it means all the integration we have to do splits into independent 1 dimensional integrals.

We begin by defining the translation, dilation and rotation operators for $\mathbb{R}^n$.

- For all $\vec{\alpha} \in \mathbb{R}^n$ we define $\mathcal{T}_{\vec{\alpha}} : L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$ by $(\mathcal{T}_{\vec{\alpha}} f)(\vec{x}) = f(\vec{x} + \vec{\alpha})$. This is a continuous linear operator with norm 1 on $L^2(\mathbb{R}^n)$.

- For all $\lambda \in \mathbb{R} \setminus \{0\}$ we define $\mathcal{D}_\lambda : L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$ by $(\mathcal{D}_\lambda f)(\vec{x}) = f(\lambda \vec{x})$. This is a continuous linear operator with norm $|\lambda|^{-\frac{n}{2}}$.

- For all $\vec{\alpha} \in \mathbb{R}^n$ we define $\mathcal{R}_{\vec{\alpha}} : L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$ by $(\mathcal{R}_{\vec{\alpha}} f)(\vec{x}) = e^{i\vec{\alpha}\cdot\vec{x}} f(\vec{x})$. This is also continuous linear operator with norm 1.

We now define the Fourier transform, more or less as before. Let Q be the unit cube $[0, 1)^n$, and D^n be all the functions in the MRA of $L^2(\mathbb{R}^n)$ generated by χ_Q. Now let $\mathcal{F}$ be defined by:

1. $\mathcal{F} : D^n \to L^2(\mathbb{R}^n)$ is linear,
2. $\mathcal{F}\mathcal{T}_{\vec{n}} = \mathcal{R}_{\vec{n}}\mathcal{F}$ where $\vec{n} \in \mathbb{Z}^n$,
3. $\mathcal{F}\mathcal{D}_\lambda = \frac{1}{|\lambda|^n}\mathcal{F}\mathcal{D}_{\frac{1}{\lambda}}$ where $\lambda = 2^n, n \in \mathbb{Z}$,
4. $\mathcal{F}(\chi_Q)(\vec{\omega})$ is continuous at $\vec{0}$ and $\mathcal{F}(\chi_Q)(\vec{0}) = 1$.

Our first task will be to derive $\mathcal{F}(\chi_Q)(\vec{\omega})$, as we did for the 1 dimensional case in Theorem 3.11. The proof is basically the same, but involves some work with a more complicated dilation equation.

Theorem 3.13. *If $\mathcal{F}$ is defined by the above rules then:*

$$\mathcal{F}(\chi_Q)(\vec{\omega}) = \prod_{k=1}^{n} \frac{1 - e^{-i\omega_k}}{i\omega_k}$$

where $\vec{\omega} = (\omega_1, \omega_2, \ldots, \omega_n)$.

Proof. Our proof of Theorem 3.11 involved looking at a dilation equation that $\chi_{[0,1)}$ satisfied. We have to find the equivalent equation for χ_Q. Examining the 1, 2 and 3 dimensional cases it becomes apparent that the equation we are looking for is:

$$\chi_Q = \sum_{\vec{r}\in\{0,1\}^n} \mathcal{D}_2 \mathcal{T}_{\vec{r}} \chi_Q$$

That is: *the unit cube is the union of the cubes of side 1/2, which have had their origins translated to each of the corners of a cube of side 1/2 at the origin.*

Now we take the transform of this and apply our dilation and translation rules to get:

$$\mathcal{F}(\chi_Q) = \sum_{\vec{r}\in\{0,1\}^n} \frac{1}{2^n} \mathcal{D}_{\frac{1}{2}} \mathcal{R}_{\vec{r}} \mathcal{F}(\chi_Q)$$

Rewriting this with f for $\mathcal{F}(\chi_Q)$ and looking at a point $\vec{\omega}$ we find:

$$f(\vec{\omega}) = \frac{1}{2^n} f\left(\frac{\vec{\omega}}{2}\right) \sum_{\vec{r}\in\{0,1\}^n} e^{-i\frac{\vec{r}\cdot\vec{\omega}}{2}}$$

Concentrating on the sum for the moment, we pluck off each component of $\vec{\omega}$. If $\vec{x} = (x_1, x_2, \ldots x_n)$ we denote $(x_2, \ldots x_n)$ by $\vec{x}'$.

$$\begin{aligned}
\sum_{\vec{r}\in\{0,1\}^n} e^{-i\frac{\vec{r}\cdot\vec{\omega}}{2}} &= \sum_{\vec{r}\in\{0,1\}^{n-1}} \left(e^{-i\frac{0\cdot\omega_1}{2}} + e^{-i\frac{1\cdot\omega_1}{2}}\right) e^{-i\frac{\vec{r}\cdot\vec{\omega}'}{2}} \\
&= \left(1 + e^{-i\frac{\omega_1}{2}}\right)\left(1 + e^{-i\frac{\omega_2}{2}}\right) \sum_{\vec{r}\in\{0,1\}^{n-2}} e^{-i\frac{\vec{r}\cdot\vec{\omega}''}{2}} \\
&= \prod_{k=1}^{n} \left(1 + e^{-i\frac{\omega_k}{2}}\right)
\end{aligned}$$

We can now substitute this expression back into our equation for $f(\vec{\omega})$:

$$\begin{aligned} f(\vec{\omega}) &= \frac{1}{2^n} f\left(\frac{\vec{\omega}}{2}\right) \prod_{k=1}^{n} \left(1 + e^{-i\frac{\omega_k}{2}}\right) \\ &= \frac{1}{2^{nM}} f\left(\frac{\vec{\omega}}{2^M}\right) \prod_{m=1}^{M} \prod_{k=1}^{n} \left(1 + e^{-i\frac{\omega_k}{2^m}}\right) \qquad \text{repeating } M \text{ times} \\ &= \frac{1}{2^{nM}} f\left(\frac{\vec{\omega}}{2^M}\right) \prod_{k=1}^{n} \prod_{m=1}^{M} \left(1 + e^{-i\frac{\omega_k}{2^m}}\right) \\ &= \frac{1}{2^{nM}} f\left(\frac{\vec{\omega}}{2^M}\right) \prod_{k=1}^{n} \sum_{r=1}^{2^M} e^{-i\frac{\omega_k}{w^M} r} \\ &= f\left(\frac{\vec{\omega}}{2^M}\right) \prod_{k=1}^{n} \frac{1}{2^M} \frac{e^{-i\omega_k} - 1}{e^{-i\frac{\omega_K}{2^M}} - 1} \end{aligned}$$

Letting $M \to \infty$, and remembering we assumed f was continuous at zero,

$$f(\vec{\omega}) = f(\vec{0}) \prod_{k=1}^{n} \frac{1 - e^{-i\omega_k}}{i\omega_k}$$

Finally using $f(\vec{0}) = 1$ gives the required result. ■

The remainder of the work to show everything is well defined, continuous, extendible, and invertible is the same as that for the one dimensional case, except for the fact that we have multidimensional integrals, which split nicely into n one dimensional integrals. To give the idea we will prove one of the results, the counterpart of Lemma 3.6.

Lemma 3.14. *If $f \in D^n$ then $\|f\|_2^2 = C\|\mathcal{F}f\|_2^2$, where $C = (2\pi)^n$.*

Proof. Keeping a close eye on Lemma 3.6 we first write f in the form:

$$f(\vec{x}) = \sum_{\vec{r}} a_{\vec{r}} \chi_Q(2^J \vec{x} - \vec{r})$$

where the sum is over some finite subset of $\mathbb{Z}^n$ and $J \in \mathbb{N}$. As in Lemma 3.6 we note that $\|f\|_2^2 = \sum |a_{\vec{r}}|^2/2^{Jn}$. Taking the Fourier transform we see:

$$\mathcal{F}(f)(\vec{\omega}) = \left(\prod_{k=1}^{n} \frac{1 - e^{-i\frac{\omega_k}{2^J}}}{i\omega_k}\right) \sum_{\vec{r}} a_{\vec{r}} e^{-i\frac{\vec{r}\cdot\vec{\omega}}{2^J}}$$

Rearranging so that all the ω_k terms are within the product:

$$\mathcal{F}(f)(\vec{\omega}) = \sum_{\vec{r}} a_{\vec{r}} \left(\prod_{k=1}^{n} \frac{1 - e^{-i\frac{\omega_k}{2^J}}}{i\omega_k} e^{-i\frac{r_k\omega_k}{2^J}} \right)$$

Now we can work out the norm, by splitting the multiple integral, and using our original calculation in Lemma 3.6:

$$\begin{aligned}
&\|\mathcal{F}f\|_2^2 \\
&= \int_{\mathbb{R}^n} \sum_{\vec{r}} a_{\vec{r}} \sum_{\vec{s}} \overline{a}_{\vec{s}} \prod_{k=1}^{n} \left| \frac{1 - e^{-i\frac{\omega_k}{2^J}}}{i\omega_k} \right|^2 e^{-i\frac{r_k\omega_k}{2^J}} e^{i\frac{s_k\omega_k}{2^J}} \, d\vec{\omega} \\
&= \sum_{\vec{r}} \sum_{\vec{s}} a_{\vec{r}} \overline{a}_{\vec{s}} \prod_{k=1}^{n} \int_{\mathbb{R}} \left| \frac{1 - e^{-i\frac{\omega_k}{2^J}}}{i\omega_k} \right|^2 e^{-i\frac{r_k\omega_k}{2^J}} e^{i\frac{s_k\omega_k}{2^J}} \, d\omega_k \\
&= \sum_{\vec{r}} \sum_{\vec{s}} a_{\vec{r}} \overline{a}_{\vec{s}} \prod_{k=1}^{n} \frac{2\pi}{2^J} \delta_{r_k s_k} \\
&= \sum_{\vec{r}} |a_{\vec{r}}|^2 \left(\frac{2\pi}{2^J} \right)^n = (2\pi)^n \|f\|_2^2
\end{aligned}$$

as required. ■

3.9 $L^p(\mathbb{R}^n)$ for other p.

Extending $\mathcal{F}$ to $L^2(\mathbb{R}^n)$ is only one direction in which we can expand — we can also make extensions in other directions. The traditional definition of the Fourier transform provides us with a definition for $\mathcal{F}$ on $L^p(\mathbb{R}^n)$ when $1 \leq p \leq 2$. We can use our definition on all these spaces too. In general $\mathcal{F} : L^p(\mathbb{R}^n) \to L^q(\mathbb{R}^n)$ with $1 \leq p \leq 2$ and $\frac{1}{p} + \frac{1}{q} = 1$. However there are no inversion results, Plancherel's Theorem or the like to look for when $p \neq 2$. This means all we are left to do is prove that $\mathcal{F}$ is bounded from and to the correct spaces (the equivalent of Lemma 3.6).

We have already dealt with the case $p = 2$, and first deal with the other extreme: the case where $p = 1$. In this case $\mathcal{F} : L^1(\mathbb{R}^n) \to L^\infty(\mathbb{R}^n)$, and we wish to show it is bounded. For the sake of simplicity we will look at $\mathbb{R}$ instead of $\mathbb{R}^n$.

Lemma 3.15. *If we take our definition of $\mathcal{F}$ on D and examine $\mathcal{F} : D \to L^\infty(\mathbb{R})$ we see $\mathcal{F}$ is a bounded operator when D is given the norm it inherits as a subset of $L^1(\mathbb{R})$.*

Proof. As in Lemma 3.6 we begin by writing f as:

$$f(x) = \sum_{k=-N}^{N} a_k \chi_{[0,1)}(2^J x - k).$$

We note that $\|f\|_1 = \frac{\sum_{k=-N}^{N} |a_k|}{2^J}$. Again, exactly as in Lemma 3.6, we calculate the Fourier transform:

$$(\mathcal{F}f)(\omega) = \frac{1 - e^{-i\frac{\omega}{2^J}}}{i\omega} \sum_{k=-N}^{N} a_k e^{-ik\frac{\omega}{2^J}}.$$

Estimating the $L^\infty(\mathbb{R})$ norm of this is now a simple matter.

$$\begin{aligned}
\|\mathcal{F}f\|_\infty &\leq \sup_{\omega\in\mathbb{R}} \left| \frac{1 - e^{-i\frac{\omega}{2^J}}}{i\omega} \sum_{k=-N}^{N} a_k e^{-ik\frac{\omega}{2^J}} \right| \\
&\leq \sup_{\omega\in\mathbb{R}} \sum_{k=-N}^{N} |a_k| \left| \frac{1 - e^{-i\frac{\omega}{2^J}}}{i\omega} e^{-ik\frac{\omega}{2^J}} \right| \\
&= \sum_{k=-N}^{N} |a_k| \sup_{\omega\in\mathbb{R}} \left| \frac{1 - e^{-i\frac{\omega}{2^J}}}{i\omega} \right| \\
&= \sum_{k=-N}^{N} |a_k| \frac{1}{2^J} \\
&= \|f\|_1
\end{aligned}$$

The second last step may be verified by basic calculus and graph sketching (see Figure 3.1 for the idea). ■

So we can now extend $\mathcal{F}$ to all of $L^1(\mathbb{R})$, in the same was as we extended $\mathcal{F}$ from D to $L^2(\mathbb{R})$ earlier.

Theorem 3.16. *We may extend $\mathcal{F} : D \to L^\infty(\mathbb{R})$ continuously to a bounded operator $\mathcal{F}_1 : L^1(\mathbb{R}) \to L^\infty(\mathbb{R})$ with norm 1 (again D is taken with $\|.\|_1$).*

Proof. We can use Lemma 3.5 to extend $\mathcal{F}$ now that we have proved Lemma 3.15. Showing that the norm is less than or equal to one also comes from Lemma 3.15. Showing it cannot be less than 1 comes from looking at the Fourier transform of $\chi_{[0,1)}$. ■

Note that we have called this new transform $\mathcal{F}_1$, because at the moment we do not know

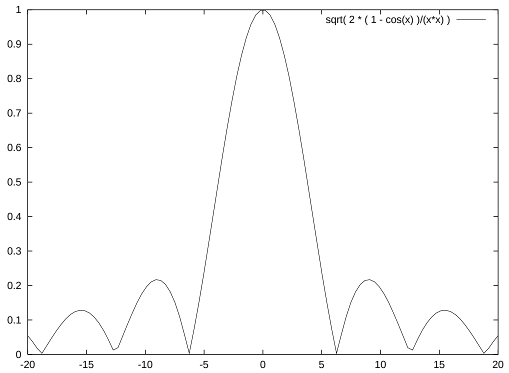


Figure 3.1: $\left|\frac{1-e^{-i\omega}}{i\omega}\right|$

how closely related to $\mathcal{F}$ it is. They must agree on D, because they are both extensions of the same function on D. It would be nice to know if they agree on all of $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$. Fortunately they do. Before we show this we will get an integral formula for $\mathcal{F}_1$.

Theorem 3.17. *If $f \in L^1(\mathbb{R})$ then*

$$\int_{-\infty}^{\infty} f(x)e^{-i\omega x}\, dx$$

agrees with $\mathcal{F}_1(f)$ almost everywhere.

Proof. Compare the this to the statement of Theorem 3.9 for the $L^2(\mathbb{R})$ case. The only change is that we no longer need to assume the integral converges almost everywhere, as $f \in L^1(\mathbb{R})$ ensures this for us.

We note that the formula above defines a continuous linear operator from $L^1(\mathbb{R})$ to $L^\infty(\mathbb{R})$. Using the first part of the proof of Theorem 3.9 we see that this same formula gives us $\mathcal{F}$ when $f \in D$. Thus, as $\mathcal{F}_1$ is the unique extension of $\mathcal{F}$, and the integral formula is another extension of $\mathcal{F}$ they must be the same. As the image of f is in $L^\infty(\mathbb{R})$, this means $\int e^{-i\omega x}\, dx$ and $\mathcal{F}_1(f)(w)$ agree for almost every ω. ∎

Theorem 3.18. *$\mathcal{F} : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ and $\mathcal{F}_1 : L^1(\mathbb{R}) \to L^\infty(\mathbb{R})$ agree on $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$.*

Proof. By comparing Theorem 3.9 and Theorem 3.17 we see that all we need to show agreement is the convergence of the integral in Theorem 3.9. However the fact the $f \in L^1(\mathbb{R})$ ensures this for us.

Using the integral formulas here might be considered "unsporting". It is possible to prove this directly, though the route is some what more bumpy.

1. If $f \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ has compact support then it is possible to choose f_n in D so that $f_n \to f$ in $L^1(\mathbb{R})$ and $f_n \to f$ in $L^2(\mathbb{R})$.

2. The definitions say $\mathcal{F}f = \lim \mathcal{F}f_n$ in $L^2(\mathbb{R})$, and $\mathcal{F}_1 f = \lim \mathcal{F}f_n$ in $L^\infty(\mathbb{R})$.

3. Now we examine $\mathcal{F}f$ and $\mathcal{F}_1 f$ in $L^2(\mathbb{R})$ and $L^\infty(\mathbb{R})$ respectively. It can be shown that these are the same almost everywhere.

4. Now we know $\mathcal{F}$ and $\mathcal{F}_1$ agree on compactly supported functions in $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$.

5. Now let g be any function in $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$. It is easy to show that $g_n = g\chi_{[-n,n)}$ converges to g in both $L^1(\mathbb{R})$ and $L^2(\mathbb{R})$.

6. Again by the definitions $\mathcal{F}g = \lim \mathcal{F}g_n$ in $L^2(\mathbb{R})$ and $\mathcal{F}_1 g = \lim \mathcal{F}_1 g_n$ in $L^\infty(\mathbb{R})$.

7. Again we can show $\mathcal{F}g = \mathcal{F}_1 g$ because we now know $\mathcal{F}g_n = \mathcal{F}_1 g_n$.

■

So, after some ado we have shown that we can define $\mathcal{F}$ on $L^1(\mathbb{R})$ and $L^2(\mathbb{R})$, and that $\mathcal{F}$ agrees on the intersection. This is a large part of the hard work involved with defining $\mathcal{F}$ on any $L^p(\mathbb{R})$ (with $1 \leq p \leq 2$) completed, due to the following lemma.

Lemma 3.19. *Let f be in $L^p(\mathbb{R})$ with $1 \leq p \leq 2$. Then we can find $f_1 \in L^1(\mathbb{R})$ and $f_2 \in L^2(\mathbb{R})$ so that $f(x) = f_1(x) + f_2(x)$.*

Proof. Define f_1 as follows and set $f_2 = f - f_1$.

$$f_1(x) = \begin{cases} f(x) & |f(x)| \geq 1 \\ 0 & \text{otherwise} \end{cases}.$$

Then using $p \geq 1$ and $|y| \geq 1$ implies $|y| \leq |y|^p$ we see that:

$$\|f_1\|_1 = \int\limits_{|f(x)| \geq 1} |f(x)|\, dx \leq \int\limits_{|f(x)| \geq 1} |f(x)|^p\, dx \leq \|f\|_p^p,$$

so $f_1 \in L^1(\mathbb{R})$. Similarly for f_2 we note that $p < 2$ and $|y| \leq 1$ implies $|y|^2 \leq |y|^p$ and so:

$$\|f_2\|_2^2 = \int\limits_{|f(x)|>1} |f(x)|^2\, dx \leq \int\limits_{|f(x)|>1} |f(x)|^p\, dx \leq \|f\|_p^p, .$$

Thus $f_2 \in L^2(\mathbb{R})$ as required. ■

Our final extension can now be performed in a straight forward manner.

Theorem 3.20. *For each $1 \leq p \leq 2$ we can extend $\mathcal{F} : D \to (L^2(\mathbb{R}) \cap L^\infty(\mathbb{R}))$ to $\mathcal{F} : L^p(\mathbb{R}) \to (L^2(\mathbb{R}) + L^\infty(\mathbb{R}))$ in a well defined manner.*

Proof. For each $f \in L^p(\mathbb{R})$ we write $f = f_1 + f_2$ where $f_1 \in L^1(\mathbb{R})$ and $f_2 \in L^2(\mathbb{R})$. Then we define:

$$\mathcal{F}f = \mathcal{F}_1 f_1 + \mathcal{F}f_2.$$

This definition is independent of the choice of f_1 and f_2, for suppose $f = g_1 + g_2$ with $g_1 \in L^1(\mathbb{R})$ and $g_2 \in L^2(\mathbb{R})$ then:

$$f_1 - g_1 = f_2 - g_2.$$

So $f_1 - g_1$ is in both $L^1(\mathbb{R})$ and $L^2(\mathbb{R})$ so Theorem 3.18 tells us $\mathcal{F}_1(f_1 - g_1) = \mathcal{F}(f_2 - g_2)$. Rearranging gives $\mathcal{F}_1(f_1) + \mathcal{F}(f_2) = \mathcal{F}_1(g_1) + \mathcal{F}(g_2)$ as required. ■

We have not said if $\mathcal{F}$ defined in this manner is bounded, and what norm we should use on the range. It seems that these calculations with the new construction are no easier than with the original construction. For this reason it would seem acceptable to use some general piece of operator interpolation theory which would tell us $\mathcal{F} : L^p(\mathbb{R}) \to L^q(\mathbb{R})$ in a bounded manner when $\frac{1}{p} + \frac{1}{q} = 1$. Chapter 5 of [16] deals with this sort of interpolation in detail.

3.10 Conclusions

This seems to be a new way of constructing the Fourier transform on $L^2(\mathbb{R}^n)$. It is quite straight forward, and doesn't need many other complicated results (the main one might be similar to Lusin's theorem, which would be used to show the dyadic step functions are dense in $L^p(\mathbb{R}^n)$).

Traditional constructions of the Fourier transform (on $L^2(\mathbb{R}^n)$) usually begin with defining the Fourier transform on some dense subset, such as $L^2(\mathbb{R}^n) \cap L^1(\mathbb{R}^n)$ or the Schwartz class $\mathcal{S}$ of C^∞ functions which decay rapidly†. These spaces are chosen because the Fourier transform will be well behaved there, for instance the integral formula is well defined on $L^2(\mathbb{R}^n) \cap L^1(\mathbb{R}^n)$, and the Fourier transform is invertible on $\mathcal{S}$ and very well behaved there. The approach we have used is similar in that we began by defining $\mathcal{F}$ on a dense subspace — the dyadic step functions D.

The definition used on these subspaces is usually the integral formula. In our MRA based construction we have used a set of desired properties to define Fourier transform, so this is one place where this new construction differs.

Once the definition has been made, the extension to all of $L^2(\mathbb{R}^n)$ is carried out in the obvious way — by proving that the transform is bounded on the chosen dense subspace and then extending it continuously. Here the MRA based method is, in spirit, the same (we also extend $\mathcal{F}$ continuously) however the proof that $\mathcal{F}$ is bounded is quite different.

The usual demonstration that $\mathcal{F}$ is bounded uses the convolution result: $\mathcal{F}(f * g) = \mathcal{F}(f)\mathcal{F}(g)$ for the Fourier transform. This is applied to f and $g(x) = \overline{f(-x)}$ to get a h with the property $\hat{h} = |\hat{f}|^2$ and after some convergence, invertibility and continuity arguments $h(0)$ is examined to discover:

$$\int |\hat{f}|^2 = \int \hat{h} = h(0) = \int f(x)g(0-x)\,dx = \int f\overline{f} = \int |f|^2$$

The convergence issues surrounding this are usually tricky and involve some actual calculation of integrals. Even the slick treatment of this subject in Chapter 1 of [16] involves some messy integration. In comparison the MRA method seems a relatively straight forward piece of contour integration.

The other main result, the invertibility of $\mathcal{F}$ on $L^2(\mathbb{R}^n)$, is usually obtained by noting that it is invertible on the dense subspace and then extending continuously. Showing that the transform is invertible on your dense subspace is not straight forward, but has usually been dealt with on the way to showing that $\mathcal{F}$ is bounded. Here the MRA construction involves a relatively simple piece of integration to show that a constructed operator is actually the inverse of $\mathcal{F}$.

The $L^1(\mathbb{R}^n)$ case, in the MRA construction, is very similar to the case $L^2(\mathbb{R}^n)$. Tradi-

†Rapidly here means that:

$$\sup \left| x^\alpha \frac{\partial^{|\beta|}}{\partial x^\beta} f(x) \right| < \infty$$

for all non-negative n-tuples α, β.

tional constructions on $L^1(\mathbb{R}^n)$ are more or less trivial, as the integral formula is as well behaved as is needed on $L^1(\mathbb{R}^n)$. The MRA construction does not seem to shed any light on the $L^p(\mathbb{R})$ case and so we resort to traditional techniques to form the Fourier transform here.

There are disadvantages to the MRA based method — it doesn't provide a nice link to the Fourier transform of tempered distributions[‡], as the dyadic step functions aren't really suitable as test functions. Neither does it easily provide access to the convolution results. The reason no convolution result is easily forthcoming seems to be because the convolution of two step functions is not a step function.

A possible way around these problems might be to use a different generating function. Something like e^{-x^2} might be quite suitable, first because it is a suitable test function for distributions, and second because it is almost its own Fourier transform. The other possible way around would be to use the integral formula, which we may as well use if the proof using it is more straight forward.

In comparison the translation, dilation and rotation properties we used to define the Fourier transform are easily derived from the integral formula, and so are not difficult to obtain in any of the traditional constructions.

It should also be possible to use wavelets instead of MRA for the basis of these constructions. This would have advantages, as we should have fewer problems with checking if things are well defined (this is as wavelets form a basis, and we will have uniqueness of representation). On the other hand higher dimensional wavelets can be more irregular than higher dimensional MRAs, so there could be a down side too.

A final note would be that it should be possible to do similar studies of other transforms (Hilbert, Laplace?), and on other spaces (Heisenberg group?). I have looked at some of these, but found that the integration is not as straight forward, and so reduces the usefulness of the method. This, naturally, is not to say that such a study is impossible or uninteresting.

[‡]The construction on $\mathcal{S}$ provides a strong link.

Chapter 4

A Uniqueness Result for the Fourier Transform

4.1 Introduction

What we now aim to find out is: are dilation, translation and rotation conditions enough to pin down the Fourier transform exactly? From the previous section it is apparent that given rules about how a linear transform behaves when composed with translation and dilation, we can make deductions about the image of functions satisfying certain dilation equations.

If we consider a linear transform $\mathcal{A} : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ with:

$$\mathcal{A}\mathcal{T}_n f = \mathcal{R}_n \mathcal{A} f \qquad \forall n \in \mathbb{Z}$$

$$\mathcal{A}\mathcal{D}_\lambda f = \frac{1}{|\lambda|}\mathcal{D}_{\frac{1}{\lambda}}\mathcal{A} f \qquad \forall \lambda = 2^n,\ n \in \mathbb{Z},$$

then by looking at the composition $\mathcal{F}^{-1}\mathcal{A}$ we would get an identity-like transform. More specifically, the solution space of any dilation equation of the form:

$$f(x) = \sum_n c_n f(2x - n)$$

would be sent to itself under $\mathcal{F}^{-1}\mathcal{A}$, because $\mathcal{F}^{-1}$ "undoes" the intertwining that $\mathcal{A}$ does.

For instance: we know $\chi_{[0,1)}$ satisfies:

$$f(x) = f(2x) + f(2x-1)$$

so $\mathcal{F}^{-1}\mathcal{A}\chi_{[0,1)}$ also satisfies it. We can conclude that $\mathcal{A}\chi_{[0,1)} = \mathcal{F}f$, where f is a solution of the above dilation equation. Now, if we know the image of some function (in this example $\mathcal{A}\chi_{[0,1)}$) we could proceed as we did for $\mathcal{F}$ in Chapter 3 to try to determine this transform on all of $L^2(\mathbb{R})$.

So looking for solutions of dilation equations, can be related to looking for transforms $\mathcal{A}$ which have certain translation and dilation properties.

It would be more accurate to write the above dilation equations in the form:

$$f = \sum_n c_n \mathcal{D}_2 \mathcal{T}_n f$$

because we will be looking for $L^2(\mathbb{R})$ solutions, which may fail to satisfy the dilation equation on some set of measure zero. We look for $L^2(\mathbb{R})$ solutions because we will want the Fourier like transforms to be linear maps on $L^2(\mathbb{R})$.

We could look for solutions in other spaces of functions. The case where the Fourier transform of the solution is continuous has been examined in some detail by those interested in wavelets. Here people are looking for compactly supported functions in $L^p(\mathbb{R})$ — Daubechies and Lagarias deal with $L^1(\mathbb{R})$ solutions of general dilation equations in [4]. When $p \geq 1$ the compactly supported $L^p(\mathbb{R})$ case is less general than the $L^1(\mathbb{R})$ case. Others have dealt with this, for example [21].

4.2 A simple start

One beginning would be to solve dilation equations that hold everywhere. In looking for solutions in $L^2(\mathbb{R})$ we will not have this option, but this search gives us a foothold.

We observe that by chopping $\chi_{[0,1)}$ at $\alpha \in (0,1)$ we get a two pieces $\chi_{[0,\alpha)}$ and $\chi_{[\alpha,1)}$ — this means $\chi_{[0,1)}$ satisfies the dilation equation:

$$\chi_{[0,1)}(x) = \chi_{[0,1)}\left(\frac{x}{\alpha}\right) + \chi_{[0,1)}\left(\frac{x-\alpha}{1-\alpha}\right)$$

for any $\alpha \in (0,1)$.

Theorem 4.1. *Let* $g : \mathbb{R} \to \mathbb{R}$ *be such that:*

$$g(x) = g\left(\frac{x}{\alpha}\right) + g\left(\frac{x-\alpha}{1-\alpha}\right)$$

$\forall \alpha \in (0,1)$. *Then* g *is a constant multiple of* $\chi_{[0,1)}$ *except at 0 and 1.*

Proof. We prove this by varying the parameter α. Setting $x = 0$:

$$\begin{aligned} g(0) &= g(0) + g\left(\frac{-\alpha}{1-\alpha}\right) \\ 0 &= g\left(\frac{-\alpha}{1-\alpha}\right) \end{aligned}$$

As α ranges over $(0,1)$, $\frac{-\alpha}{1-\alpha}$ ranges over $(-\infty, 0)$, thus g is supported on $[0,\infty)$. Setting $x = 1$ leads to a similar cancelation giving:

$$0 = g\left(\frac{1}{\alpha}\right)$$

and as α ranges over $(0,1)$, $1/\alpha$ goes from 1 to ∞. So now we know g is supported on $[0,1]$. Now to get the values in the middle we set $x = 1/2$:

$$g\left(\frac{1}{2}\right) = g\left(\frac{1}{2\alpha}\right) + g\left(\frac{1-2\alpha}{2(1-\alpha)}\right)$$

Now when $\alpha \in (0, \frac{1}{2})$, $1/2\alpha > 1$ and $(1-2\alpha)/2(1-\alpha)$ ranges over $(0, \frac{1}{2})$. Similarly if $\alpha \in (\frac{1}{2}, 1)$, $(1-2\alpha)/2(1-\alpha) < 0$ and $1/2\alpha$ ranges over $(\frac{1}{2}, 1)$. Filling in to the above we find that $g(x) = g(1/2)$ for $x \in (0,1)$.

We also observe that by putting $x = \alpha = 1/2$ gives that $g(0) + g(1) = g(1/2)$.

■

Remark 4.1. It is worth noting that we have used a lot of dilation equations in the proof of this theorem, one for each $\alpha \in (0,1)$ — an uncountable number. This is the theorem's undoing in $L^2(\mathbb{R})$.

4.3 A useful counterexample

The unlikely looking function,

$$\log\left|1-\frac{1}{x}\right|,$$

is actually closely related to $\chi_{[0,1)}$. It is the Hilbert transform* of $\chi_{[0,1)}$, and so has Fourier transform:

$$-i\,\mathrm{sign}(\omega)\frac{1-e^{-i\omega}}{i\omega}.$$

While trying to weaken the definition of the Fourier transform in Section 3.5 we came up with this type of function as a possible contender for the image of $\chi_{[0,1)}$. To rule it out we introduced the continuity at 0 clause. Another reason that this function is interesting is because it only narrowly misses being in $L^1(\mathbb{R})$.

Lemma 4.2. *Let $\phi(x)$ be given by:*

$$\phi(x)=\log\left|1-\frac{1}{x}\right|.$$

Then $\phi\in L^p(\mathbb{R})$ for $1<p<\infty$.

Proof. A sketch of ϕ (Figure 4.1), shows we have two things to worry about. The tails of the function, and the two points where it blows up. It does also show that the function is symmetric (under reflection in $(1/2,0)$), so at least we only have to do one tail and either of the bad points.

The tails are easy to deal with. Looking at

$$\frac{d\phi}{dx}=\frac{1}{x^2-x}$$

we see that $\phi(x)$ behaves like $-\frac{1}{x}$ in the limit as $|x|\to\infty$. To make this precise we compare

*The Hilbert transform $\mathcal{H}$ of a function f is usually given to be

$$(\mathcal{H}f)(x)=\frac{1}{\pi}\int\frac{f(t)}{x-t}\,dt.$$

This corresponds to multiplying the Fourier transform by $-i\,\mathrm{sign}(\omega)$. The Hilbert transform is mentioned in most books dealing with the theory of Fourier transforms, for example [16].

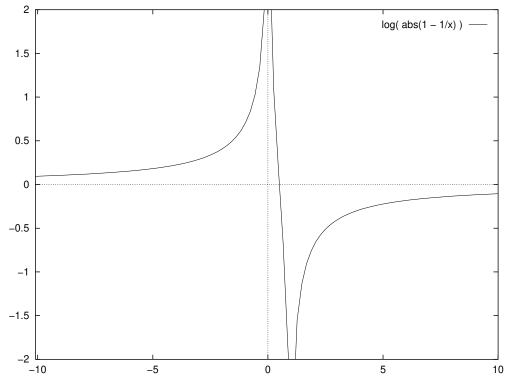


Figure 4.1: $\phi(x)$

$-\frac{1}{x}$ with $\phi(x)$ on $(2,\infty)$. Since the limit of both is zero we can write (for $a \geq 2$):

$$\begin{aligned}
\phi(a) + \frac{1}{a} &= \int_{\infty}^{a} \frac{d}{dx}\left(\phi(x) + \frac{1}{x}\right) dx \\
&= \int_{\infty}^{a} \frac{1}{x^2 - x} - \frac{1}{x^2}\, dx \\
&= \int_{\infty}^{a} \frac{1}{x^2(x-1)}\, dx \\
\left|\phi(a) + \frac{1}{a}\right| &= \int_{a}^{\infty} \frac{1}{x^2(x-1)}\, dx \\
&\leq \int_{a}^{\infty} \frac{1}{(x-1)^3}\, dx \\
&= \frac{1}{2(a-1)^2}
\end{aligned}$$

If we look at this when x is large (say $|x| > 2$) we see the tails of $\frac{1}{a}$ and $\frac{1}{2(a-1)^2}$ are in $L^p(\mathbb{R})$ when $p > 1$, so the tails of ϕ are also in $L^p(\mathbb{R})$.

Now we turn to the singularity at 0. We try to see how bad it is in the same way we might examine a pole, by calculating:

$$\lim_{x \to 0} x^n \phi(x)$$

for $n > 0$. We will have to use L'Hôpital's rule here.

$$
\begin{aligned}
\lim_{x\to 0} \frac{\log\left|1-\frac{1}{x}\right|}{x^{-n}} &= \lim_{x\to 0} \frac{\frac{1}{x^2-x}}{-nx^{-(n+1)}} \\
&= \lim_{x\to 0} \frac{x^{n+1}}{-n(x^2-x)} \\
&= \lim_{x\to 0} \frac{x^{n}}{-n(x-1)} \\
&= 0.
\end{aligned}
$$

We may choose n so that $np < 1$, as $p < \infty$. So $\forall \epsilon > 0\ \exists\, \delta > 0$ so that when $|x| < \delta$ we know:

$$
\begin{aligned}
\left|x^n \log\left|1-\frac{1}{x}\right|\right| &< \epsilon \\
\left|\log\left|1-\frac{1}{x}\right|\right| &< \frac{\epsilon}{|x|^n} \\
\int_{-\delta}^{\delta} \left|\log\left|1-\frac{1}{x}\right|\right|^p dx &< \int_{-\delta}^{\delta} \frac{\epsilon^p}{|x|^{np}}\, dx \\
&< \infty.
\end{aligned}
$$

So now we know the tails and singularities contribute only a finite amount to the $L^p(\mathbb{R})$ norm, and the remaining bits are bounded and have compact support, so ϕ has finite $L^p(\mathbb{R})$ norm as required.

■

This example leads in to three useful pieces of information.

Remark 4.2. Examining its dilation properties we find ϕ satisfies all the criteria of Theorem 4.1, except when ϕ is evaluated at 0 and 1. At $x \neq 0, \alpha, 1$:

$$
\begin{aligned}
\phi\left(\frac{x}{\alpha}\right) + \phi\left(\frac{x-\alpha}{1-\alpha}\right) &= \log\left|1-\frac{\alpha}{x}\right| + \log\left|1-\frac{1-\alpha}{x-\alpha}\right| \\
&= \log\left|\left(\frac{x-\alpha}{x}\right)\left(\frac{x-\alpha-1+\alpha}{x-\alpha}\right)\right| \\
&= \log\left|1-\frac{1}{x}\right| = \phi(x)
\end{aligned}
$$

The fact that the set over which α ranges has non-zero measure causes a problem. When we consider all the equations together, the set where at least one of the dilation equations fails to hold is:

$$\bigcup_{\alpha\in(0,1)} \{0, \alpha, 1\} = [0, 1],$$

which has non-zero measure. This is why we cannot try to use Theorem 4.1 when we want solutions which hold almost everywhere.

Remark 4.3. The symmetries of ϕ are slightly different to those of $\chi_{[0,1)}$. About the line $x = 1/2$, ϕ is odd but $\chi_{[0,1)}$ is even. Properties like odd and even can be expressed as dilation type equations with negative scale factors. In this case:

$$\chi_{[0,1)}(x) = \chi_{[0,1)}(-x + 1)$$

$$\phi(x) = -\phi(-x + 1)$$

Remark 4.4. The fact that $\mathcal{F}\phi$ is $\mathcal{F}\chi_{[0,1)}$ multiplied by some relatively uncomplicated function is quite interesting. It will turn out to be profitable to look at the ratio of the Fourier transforms of solutions to a given dilation equation.

4.4 How many solutions?

So we have found that $\chi_{[0,1)}$ and ϕ both satisfy the simple dilation equation:

$$f(x) = f(2x) + f(2x - 1)$$

We might well suspect that the solution space of this equation is two dimensional, as it has two terms, a dilation factor of two and its coefficients sum to two. If this were the case all we would need to pin down the Fourier transform would be:

$$\mathcal{A}\mathcal{T}_n f = \mathcal{R}_n \mathcal{A} f \qquad \forall n \in \mathbb{Z}$$

$$\mathcal{A}\mathcal{D}_\lambda f = \frac{1}{|\lambda|}\mathcal{D}_{\frac{1}{\lambda}}\mathcal{A} f \qquad \forall\lambda = \pm 2^n,\ n \in \mathbb{Z}$$

because using the negative scales we could rule out ϕ by looking at the transformed versions of the dilation equations in Remark 4.3.

Unfortunately, the solution space is not 2 dimensional, it is quite a lot bigger. To find some more solutions we must go back to the way we found ϕ in the first place — by looking at conditions on its transform.

Lemma 4.3. *Let $\mathcal{A} : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ be a linear transform which satisfies:*

$$\mathcal{A}\mathcal{T}_n f = \mathcal{R}_n \mathcal{A} f \qquad \forall n \in \mathbb{Z}$$

$$\mathcal{A}\mathcal{D}_\lambda f = \frac{1}{|\lambda|}\mathcal{D}_{\frac{1}{\lambda}} \mathcal{A} f \qquad \forall \lambda = \pm 2^n,\ n \in \mathbb{Z}$$

If we denote $\mathcal{A}\chi_{[0,1)}$ by ψ, then ψ satisfies:

$$\psi(x) = e^{-ix}\psi(-x)$$

and:

$$\psi(x) = \frac{1}{2}\frac{1-e^{-ix}}{1-e^{\frac{-ix}{2}}}\psi\left(\frac{x}{2}\right)$$

for almost all $x \in \mathbb{R}$.

Proof. To get the first condition we rearrange the dilation equation:

$$\begin{aligned}
\chi_{[0,1)}(x) &= \chi_{[0,1)}(-x+1) \\
\chi_{[0,1)} &= \mathcal{D}_{-1}\mathcal{T}_1\chi_{[0,1)} \\
\mathcal{A}\chi_{[0,1)} &= \mathcal{A}\mathcal{D}_{-1}\mathcal{T}_1\chi_{[0,1)} \\
&= \mathcal{D}_{-1}\mathcal{A}\mathcal{T}_1\chi_{[0,1)} \\
&= \mathcal{D}_{-1}\mathcal{R}_1\mathcal{A}\chi_{[0,1)} \\
\psi(x) &= \mathcal{D}_{-1}\mathcal{R}_1\psi(x) \quad \text{a.e.} \quad x \in \mathbb{R} \\
\psi(x) &= \mathcal{D}_{-1}e^{ix}\psi(x) \quad \text{a.e.} \quad x \in \mathbb{R} \\
\psi(x) &= e^{-ix}\psi(-x) \quad \text{a.e.} \quad x \in \mathbb{R}
\end{aligned}$$

To get the second we rearrange the dilation equation:

$$\begin{aligned}
\chi_{[0,1)}(x) &= \chi_{[0,1)}(2x) + \chi_{[0,1)}(2x-1) \\
\chi_{[0,1)} &= \mathcal{D}_2\chi_{[0,1)} + \mathcal{D}_2\mathcal{T}_{-1}\chi_{[0,1)} \\
\mathcal{A}\chi_{[0,1)} &= \mathcal{A}\mathcal{D}_2\chi_{[0,1)} + \mathcal{A}\mathcal{D}_2\mathcal{T}_{-1}\chi_{[0,1)} \\
\psi &= \frac{1}{2}\mathcal{D}_{\frac{1}{2}}\psi + \frac{1}{2}\mathcal{D}_{\frac{1}{2}}\mathcal{A}\mathcal{T}_{-1}\chi_{[0,1)} \\
\psi &= \frac{1}{2}\mathcal{D}_{\frac{1}{2}}\psi + \frac{1}{2}\mathcal{D}_{\frac{1}{2}}\mathcal{R}_{-1}\psi \\
\psi(x) &= \frac{1}{2}\psi\left(\frac{x}{2}\right) + \frac{1}{2}e^{-i\frac{x}{2}}\psi\left(\frac{x}{2}\right) \quad \text{a.e.} \quad x \in \mathbb{R} \\
\psi(x) &= \frac{1}{2}\left(1 + e^{-i\frac{x}{2}}\right)\psi\left(\frac{x}{2}\right) \quad \text{a.e.} \quad x \in \mathbb{R} \\
\psi(x) &= \frac{1}{2}\frac{1-e^{-ix}}{1-e^{-i\frac{x}{2}}}\psi\left(\frac{x}{2}\right) \quad \text{a.e.} \quad x \in \mathbb{R}
\end{aligned}$$

■

Now to some extent we can see where we are going. Having a rule for dilation by 2^n tells us the relation between $\psi(x)$ and $\psi(2^n x)$. Having a rule for dilation by -1 gives us a relation between $\psi(x)$ and $\psi(-x)$.

It would even seem that these are the only constraints. We can use this to get some results about solving general dilation equations. Indeed suppose that we have some solution f_0 to a dilation equation. When we take the Fourier† transform of this dilation equation suppose we get:

$$\hat{f}(x) = P(x)\hat{f}\left(\frac{x}{2}\right)$$

Naturally we know $\hat{f}_0$ satisfies this. If we pick some function g so that $g(x) = g(2x)$ for all $x \in \mathbb{R}$, then $g(x)\hat{f}_0(x)$ will also satisfy this transformed dilation equation. If this $g(x)\hat{f}_0(x)$ is in $L^2(\mathbb{R})$ then we can take the inverse transform and we have a new solution to the dilation equation.

This tells us that there are lots of solutions of our dilation equations. Daubechies and Lagarias do write about this type of solution in [4], but the fact that they are looking for $f \in L^1(\mathbb{R})$ means $\hat{f}$ is continuous. For this reason they reject these solutions in many cases (the idea is that they only accept these solutions where either g is constant, or $\hat{f}_0(x) \to 0$

†$\mathcal{F}$ would be a suitable $\mathcal{A}$ in Lemma 4.3

as $x \to 0$ and $\hat{f}_0$ and g are continuous).

They construct suitable g by choosing two period 1 functions g_+ and g_- and then define g by:

$$g(x) = \begin{cases} g_+ (\log_2 |x|) & x > 0 \\ g_- (\log_2 |x|) & x < 0 \\ 0 & x = 0 \end{cases}.$$

The following lemma just makes a concrete case for our dilation equation:

$$f(x) = f(2x) + f(2x-1).$$

Lemma 4.4. *Let $f_0 = \chi_{[0,1)}$ and let E be a measurable subset of $(1,2)$. Define g on $(1,2)$ so that $g = \chi_E$, and extend g uniquely so that $g(x) = g(-x)$ and $g(x) = g(2x)$. Then:*

$$f(x) = \mathcal{F}^{-1} \left(g \mathcal{F} f_0 \right)(x)$$

is a solution to $f(x) = f(2x) + f(2x-1)$.

Proof. Since f_0 is in $L^2(\mathbb{R})$, $\mathcal{F} f_0 \in L^2(\mathbb{R})$. Let $F = \bigcup_{n \in \mathbb{Z}} \pm 2^n E$ then:

$$\begin{aligned} \int_{\mathbb{R}} |g \mathcal{F} f_0|^2 \, dx &= \int_F |\mathcal{F} f_0|^2 \, dx \\ &\leq \int_{\mathbb{R}} |\mathcal{F} f_0|^2 \, dx \\ &< \infty \end{aligned}$$

So $g\mathcal{F} f_0$ is in $L^2(\mathbb{R})$, so we can find $\mathcal{F}^{-1}$ of it.

Verifying it is a solution of the dilation equation is just a matter of working through the algebra. ∎

This more or less puts the last nail in the coffin of any attempt to try to uniquely determine the Fourier transform using dilations by only powers of 2. Even with dilations by negative factors we will still have a huge amount of freedom. Our only hope of success is to introduce dilations by more scales.

4.5 Some more dilation equations

While proving Theorem 4.1 we used the fact that $\chi_{[0,1)}$ was a solution of

$$f(x) = f\left(\frac{x}{\alpha}\right) + f\left(\frac{x-\alpha}{1-\alpha}\right)$$

for all α in $(0,1)$. Let's see what happens if we have a transform $\mathcal{A}$ for which the dilation property held for powers of both α and $1-\alpha$.

$$f = \mathcal{D}_{\frac{1}{\alpha}} f + \mathcal{T}_{-\alpha}\mathcal{D}_{\frac{1}{1-\alpha}} f$$

Note that we are going to need to use translation by many α now, and not just integers.

$$\begin{aligned}
\mathcal{A}f &= \mathcal{A}\mathcal{D}_{\frac{1}{\alpha}} f + \mathcal{A}\mathcal{T}_{-\alpha}\mathcal{D}_{\frac{1}{1-\alpha}} f \\
\mathcal{A}f &= \alpha\mathcal{D}_{\alpha}\mathcal{A}f + \mathcal{R}_{-\alpha}\mathcal{A}\mathcal{D}_{\frac{1}{1-\alpha}} f \\
\mathcal{A}f &= \alpha\mathcal{D}_{\alpha}\mathcal{A}f + (1-\alpha)\mathcal{R}_{-\alpha}\mathcal{D}_{1-\alpha}\mathcal{A}f \\
(\mathcal{A}f)(x) &= \alpha(\mathcal{A}f)(\alpha x) + (1-\alpha)e^{-i\alpha x}(\mathcal{A}f)((1-\alpha)x) \quad \text{a.e.} \quad x \in \mathbb{R}
\end{aligned}$$

The first problem with this equation is that it is a three scale equation: 1, α and $1-\alpha$ — this means the equation can not easily be iterated. The equation that gave us the relationship between x and $2x$ was a two scale equation, and this made it ideal for iteration. The second problem is that we don't have a good set to choose α from — if we take all the $\alpha \in (0,1)$ then we have a problem where we take the union of too many sets of measure 0, and end up with a set of positive measure. We could try $\alpha \in \mathbb{Q} \cap (0,1)$ — but in Theorem 4.1 that wouldn't be good enough without an improved argument.

So we want a new set of dilation equations with the following properties:

- Each one should only contain 2 scales.
- There should be at most a countable number of them.
- They should be simple/regular enough that we can apply our translation and dilation rules and see what happens.

Fortunately the next most obvious way of chopping up $\chi_{[0,1)}$ provides us with exactly what we need. All we do is chop $\chi_{[0,1)}$ into n equal parts ($n \in \mathbb{N}$).

$$\begin{aligned}\chi_{[0,1)}(x) &= \chi_{[0,\frac{1}{n})}(x) + \chi_{[\frac{1}{n},\frac{2}{n})}(x) + \cdots + \chi_{[\frac{n-1}{n},1)}(x) \\ &= \chi_{[0,1)}(nx) + \chi_{[0,1)}(nx-1) + \cdots + \chi_{[0,1)}(nx-n+1)\end{aligned}$$

Now we can prove a slightly more general version of Lemma 4.3.

Lemma 4.5. *Suppose f satisfies:*

$$f(x) = \sum_{r=0}^{n-1} f(nx-r)$$

for some $n \in \mathbb{N}$ and $\mathcal{A}$ is a linear transform with

$$\mathcal{A}\mathcal{T}_m f = \mathcal{R}_m \mathcal{A} f \qquad \forall m \in \mathbb{Z}$$

$$\mathcal{A}\mathcal{D}_\lambda f = \frac{1}{|\lambda|}\mathcal{D}_{\frac{1}{\lambda}}\mathcal{A} f \qquad \forall \lambda = \pm n^m,\ m \in \mathbb{Z}$$

then $\mathcal{A}f$ denoted by $\tilde{f}$ satisfies:

$$\tilde{f}(x) = \frac{1}{n}\frac{1-e^{-ix}}{1-e^{-i\frac{x}{n}}}\tilde{f}\left(\frac{x}{n}\right)$$

for almost all $x \in \mathbb{R}$.

Proof. The proof is completely analogous to that of Lemma 4.3. ■

We can now see what effect having a rule for dilation by n has — it relates $\tilde{f}(x)$ to $\tilde{f}(x/n)$. The next lemma shows that in $L^p(\mathbb{R})$ the fact that this relation is true only almost everywhere is not so important because we can redefine the function on a set of measure zero so the relation between $\tilde{f}(x)$ and $\tilde{f}(x/n)$ is true everywhere.

Lemma 4.6. *If $f \in L^p(\mathbb{R})$ satisfies a finite non-zero dilation equation of scale $\alpha > 1$, and $\mathcal{A} : L^p(\mathbb{R}) \to L^q(\mathbb{R})$ has the dilation property for scale α and the translation property for integers, then knowing $\tilde{f} = \mathcal{A}f$ on $[\alpha^n, \alpha^{n+1})$ determines $\tilde{f}$ in $L^q(\mathbb{R}^+)$.*

Proof. We have supposed that f satisfies a dilation equation:

$$f = \sum_n c_n \mathcal{D}_\alpha \mathcal{T}_{-n} f$$

Applying $\mathcal{A}$ and using the translation and dilation rules:

$$\tilde{f} = \left(\frac{1}{\alpha} \sum_n c_n e^{-i\frac{xn}{\alpha}} \right) \mathcal{D}_{\frac{1}{\alpha}} \tilde{f}.$$

Or:

$$\tilde{f}(x) = P(x) \tilde{f} \left(\frac{x}{\alpha} \right) \quad \text{a.e.} \quad x \in \mathbb{R}.$$

$P(x)$ is just a trigonometric polynomial and so is analytic, which means either P is identically 0 or P has a countable number of zeros. As the c_n are not all zero P can not be 0, so we conclude $P(x)$ has a countable number of zeros.

If all things were well behaved we could iterate our relation to get (for n in $\mathbb{N}$):

$$\begin{aligned} \tilde{f}\left(\frac{x}{\alpha^n}\right) &= \frac{\tilde{f}(x)}{\prod_{r=0}^{n-1} P\left(\frac{x}{\alpha^r}\right)} \\ \tilde{f}\left(x\alpha^n\right) &= \prod_{r=0}^{n-1} P\left(x\alpha^r\right) \tilde{f}(x). \end{aligned}$$

Thus providing $P(\alpha^n x)$ is nonzero and the original relation holds for $\alpha^n x$ then given $\tilde{f}(x)$ we may find $\tilde{f}(\alpha^n x) \forall n \in \mathbb{Z}$. So if we know $\tilde{f}$ on $[\alpha^n, \alpha^{n+1})$ and we want to know $\tilde{f}$ at $x \in \mathbb{R}^+$ we simply multiply x by some power of α until it lies in our interval $[\alpha^n, \alpha^{n+1})$ — providing we avoid the bad points.

However, we can show that the set of these bad points is of measure zero. Let N be the set of points where our original relation does not hold. Then the set of all bad points is:

$$\begin{aligned} M &= \left\{ x \in \mathbb{R}^+ : P(\alpha^n x) = 0 \text{ or } \alpha^n x \in N \text{ for some } n \in \mathbb{Z} \right\} \\ &= \{ \alpha^n x : P(x) = 0 \text{ and } n \in \mathbb{Z} \} \cup \{ \alpha^n x : x \in N \text{ and } n \in \mathbb{Z} \} \\ &= \bigcup_{n \in \mathbb{Z}} \{ \alpha^n x : P(x) = 0 \} \cup \bigcup_{n \in \mathbb{Z}} \{ \alpha^n x : x \in N \}. \end{aligned}$$

But this is the union of two countable unions of sets of measure zero, and so has measure zero. Thus we can set $\tilde{f}$ to be 0 on M without changing anything in $L^q(\mathbb{R})$, and the relation will then hold everywhere (as the relation will always hold if both $\tilde{f}(x)$ and $\tilde{f}(x/\alpha)$ are zero). ■

Remark 4.5. We should note that $[0, \epsilon]$ always contains an interval of the form $[\alpha^n, \alpha^{n+1})$ for any $\epsilon > 0$. This means knowing $\tilde{f}$ on $[0, \epsilon]$ is enough to determine $\tilde{f}$ on the whole positive side.

Remark 4.6. We proved the lemma for $\alpha > 1$, but the situation is essentially the same for scales $0 < \alpha < 1$. Also, having a negative scale and related non-zero dilation equation lets us determine $\tilde{f}$ on $\mathbb{R}^-$ from $\tilde{f}$ on $\mathbb{R}^+$.

We can now see having a dilation rule and equation of scale n has more or less the same effect as knowing one of scale 2. We know that allowing the scale to be 2 is not enough, so for the moment let's go to the other extreme and allow all $n \in \mathbb{Z}$.

4.6 Pinning it down

We are now ready to present a way of pinning the Fourier transform down. To prove this theorem we will need the following result:

Lemma 4.7. *If $g \in L^1(\mathbb{R})$, and if*

$$f(x) = \int_{-\infty}^{x} g(t)\,dt \qquad x \in (-\infty, \infty)$$

then f is continuous and

$$\frac{d}{dx} f(x) = g(x) \qquad a.e. \quad x \in \mathbb{R}$$

Proof. This result is proved in [14], Theorem 8.17 page 176. ∎

Theorem 4.8. *Let $\mathcal{A} : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ be a bounded linear transform which satisfies:*

$$\mathcal{A}\mathcal{T}_n f = \mathcal{R}_n \mathcal{A} f \qquad \forall n \in \mathbb{Z},$$

$$\mathcal{A}\mathcal{D}_\lambda f = \frac{1}{|\lambda|}\mathcal{D}_{\frac{1}{\lambda}} \mathcal{A} f \qquad \forall \lambda \in \mathbb{Z} \setminus \{0\}.$$

If we denote $\mathcal{A}\chi_{[0,1)}$ by $\tilde{f}$, then $\tilde{f}$ is a constant multiple of $\mathcal{F}\chi_{[0,1)}$ almost everywhere.

Proof. First let $\tilde{f}_0$ be $\mathcal{F}\chi_{[0,1)}$, then from Lemma 4.5 we know $\tilde{f}$ and $\tilde{f}_0$ satisfy:

$$\tilde{f} = \frac{1}{n}\frac{1 - e^{-ix}}{1 - e^{-i\frac{x}{n}}} \mathcal{D}_{\frac{1}{n}} \tilde{f}$$

for n in $\mathbb{Z} \setminus \{0\}$. Or writing $\frac{1}{n}\frac{1-e^{-ix}}{1-e^{-i\frac{x}{n}}}$ as $P_n(x)$:

$$\tilde{f} = P_n(x)\mathcal{D}_{\frac{1}{n}}\tilde{f}$$

Note that $P_n(x)$ is non-zero on $[0, 2\pi)$, and is bounded away from zero on $[0, \pi]$ (for instance). The same can be said for $\tilde{f}_0(x) = \frac{1-e^{-ix}}{ix}$, so it is safe to divide by both $\tilde{f}_0$ and $P_n(x)\mathcal{D}_{\frac{1}{n}}\tilde{f}_0$ when x is in $[0, \pi]$. Doing this we get:

$$\frac{\tilde{f}}{\tilde{f}_0} = \frac{P_n(x)\mathcal{D}_{\frac{1}{n}}\tilde{f}}{P_n(x)\mathcal{D}_{\frac{1}{n}}\tilde{f}_0} = \mathcal{D}_{\frac{1}{n}}\frac{\tilde{f}}{\tilde{f}_0},$$

as long as both sides are evaluated in $[0, \pi]$. We rewrite this in terms of $g = \tilde{f}/\tilde{f}_0$:

$$g = \mathcal{D}_{\frac{1}{n}}g,$$

again as long as both sides are evaluated in $[0, \pi]$.

We can chain some of these results together, and get a similar results for $\alpha \in \mathbb{Q} \setminus \{0\}$, instead of just $n \in \mathbb{Z} \setminus \{0\}$. Indeed suppose $\alpha = n/m$ with $n, m \in \mathbb{Z}$ and $m \neq 0$ then:

$$\mathcal{D}_{\frac{n}{m}}g = \mathcal{D}_{\frac{1}{m}}\mathcal{D}_n g = \mathcal{D}_{\frac{1}{m}}g = g$$

as long as nx/m, x/m and x are all in $[0, \pi]$ (and if x is in the correct range x/m certainly is). So $g = \mathcal{D}_\alpha g$ in our range.

Now we define $G(x) = \int_0^x g(x')\,dx'$. Remember that $\tilde{f}$ is in $L^2(\mathbb{R})$ which implies that $\tilde{f}\Big|_{[0,\pi]}$ is in $L^1(\mathbb{R})$. Combining this with the fact that $\tilde{f}_0$ is bounded away from zero on this interval we can conclude that $g|_{[0,\pi]}$ is in $L^1(\mathbb{R})$ also[‡]. So using Lemma 4.7 we conclude $G(x)$ is continuous on $[0, \pi]$.

[‡]What we really mean by $g|_{[0,\pi]} \in L^1(\mathbb{R})$ is that $g\chi_{[0,\pi]} \in L^1(\mathbb{R})$

But we can actually work out $G(x)$ on the rationals. Let α be in $\mathbb{Q}$:

$$\begin{aligned} G(\alpha) &= \int_0^\alpha g(x')\,dx' \\ &= \int_0^\alpha D_{\frac{1}{\alpha}} g(x')\,dx' \\ &= \int_0^\alpha g\left(\frac{x'}{\alpha}\right)\,dx' \\ &= \alpha \int_0^1 g(y)dy \\ &= \alpha G(1) \end{aligned}$$

So, as $\mathbb{Q}$ is dense, and $G(\alpha) = \alpha G(1)$ for $\alpha \in \mathbb{Q} \cap [0, \pi]$ and G continuous on $[0, \pi]$ we can say $G(x) = xG(1)$ for all $x \in [0, \pi]$.

Now we apply the second part of Lemma 4.7 to conclude that $G' = g$ almost everywhere on $[0, \pi]$. So:

$$\begin{aligned} g(x) &= G(1) && \text{a.e. } x \in [0, \pi] \\ \tilde{f}(x) &= G(1)\tilde{f}_0(x) && \text{a.e. } x \in [0, \pi] \\ \tilde{f}(x) &= G(1)\tilde{f}_0(x) && \text{a.e. } x \in \mathbb{R}^+ \quad \text{by Lemma 4.6} \end{aligned}$$

By using negative scales and the negative scale dilation equation in Remark 4.3 we also find that $\tilde{f}(x) = G(1)\tilde{f}_0(x)$ for almost all negative x. Thus $\tilde{f} = G(1)\tilde{f}_0$ in $L^2(\mathbb{R})$.

∎

Remark 4.7. Examining the properties of the set of dilations ($\mathbb{Z} \setminus \{0\}$) which were used in the proof, we see the only property we used was that the multiplicative group generated by the set was dense in $\mathbb{R}$. We don't actually need that many scales in our set of dilations to get this. Consider for instance $\{-1, 2, 3\}$. The multiplicative group generated by this is:

$$S = \{\pm 2^n 3^m : n, m \in \mathbb{Z}\}$$

We can examine the positive part of this set this by taking $\log_2$, so we get $\{n + m\theta : n, m \in \mathbb{Z}\}$, where θ is $\log_2 3$. As $\log_2 : \mathbb{R}^+ \to \mathbb{R}$ is a homeomorphism, the original set is dense iff this one is. But this new set is clearly invariant under translation by integers, so we just need to decide if $\{m\theta \mod 1 : m \in \mathbb{Z}\}$ is dense in $[0, 1]$. But θ is irrational therefore the set is dense in $[0, 1]$, and so the original set S is dense in $\mathbb{R}$.

We can see from this that when the set of dilations contains two scales which are not rational powers of one another we get a set dense in $\mathbb{R}^+$. Including a negative scale then gives us a set dense in all of $\mathbb{R}$.

With this remark in mind, we could use our last theorem to show the following, which really answers the question asked at the start of the section.

Theorem 4.9. *Let $\mathcal{A} : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ be a bounded transform which satisfies:*

1. *$\mathcal{A}$ is linear.*
2. *$\mathcal{A}\mathcal{T}_n f = \mathcal{R}_n \mathcal{A} f \qquad \forall n \in \mathbb{Z}$.*
3. *$\mathcal{A}\mathcal{D}_\lambda f = \frac{1}{|\lambda|}\mathcal{D}_{\frac{1}{\lambda}}\mathcal{A} f \qquad \forall \lambda \in Z \subset \mathbb{Z} \setminus \{0\}$*
4. *The multiplicative group S generated by Z is dense in $\mathbb{R}$.*

Then $\mathcal{A}$ is a constant multiple of the Fourier transform.

Proof. By using essentially the same proof as in Theorem 4.8 we show that $\mathcal{A}\chi_{[0,1)}$ is a constant multiple of $\mathcal{F}\chi_{[0,1)}$. Then using the MRA construction of the Fourier transform from Chapter 3 we show that $\mathcal{A}$ must be that multiple of the Fourier transform. ■

We could actually recast this style of proof into a uniqueness result for simultaneous dilation equations.

Theorem 4.10. *Say f and g are $L^2(\mathbb{R})$ solutions to the integer scale dilation equations:*

$$f(x) = \sum_n c_{m\,n} f(\alpha_m x - n)$$

($\alpha_m \in \mathbb{Z}$) for $m = 0, 1, 2, \ldots N$. Further suppose that:

$$S = \{\alpha_0^{n_0}\alpha_1^{n_1}\ldots\alpha_N^{n_N} : n_0, n_1 \ldots n_N \in \mathbb{Z}\}$$

is dense in $\mathbb{R}$ and that $\hat{f}(\omega) = (\mathcal{F}f)(\omega)$ is bounded away from zero on $[0, \epsilon]$ for some $\epsilon > 0$, then g is a constant multiple of f.

4.7 Conclusion

We have seen that a single dilation equation may have many solutions in $L^2(\mathbb{R})$. With further restrictions we can actually get the solution to be unique up to multiplication by a constant. The two restrictions we have encountered are:

- The function has compact support, and so is in $L^1(\mathbb{R})$, and has a continuous Fourier transform.
- The function solves a set of dilation equations of varying scales, which generate a multiplicative group that is dense in the reals.

In Chapter 3 we used the first of these restrictions to "weaken" our definition of the Fourier transform. In this chapter we used the second to give a uniqueness result: we can now pin down the Fourier transform as the only bounded transform with certain translation and dilation properties.

It should be possible to generalise this proof to $L^2(\mathbb{R}^n)$, but this would require more complicated dilation structures and an $\mathbb{R}^n$ version of Theorem 4.7. This more complex dilation would have to allow independent dilation in each direction in $\mathbb{R}^n$.

We have still said only a little about producing solutions to a dilation equation. We have, however, learned to produce new solutions from one which we have been given.

Chapter 5

Future Work

As Shelley claims Ozymandias said:

> Look on my works, ye Mighty, and despair!

Which, as intended by Shelley, can be interpreted in several ways. The work presented here is far from earth shattering (and perhaps *sufficiently* far from earth shattering that it *would* make the mighty despair), but the aspect to be focused upon in this final section is the fact that learning more only serves to highlight how little we really know. This work has almost certainly raised more questions than it has answered in the author's mind.

Even Chapter 1 and Chapter 2 leave some interesting questions. The Haar MRA and band limited MRA are in some senses duals of one another — g for the Haar MRA is the characteristic function of an interval as is $\hat{g}$ for the band limited MRA. This dual structure of the MRA may warrant further investigation.

The examples in Chapter 1 which use music to demonstrate the usefulness of the CWT raise the possibility of using the CWT to automatically produce sheet music from a recording. This could be an extremely rewarding project requiring many diverse skills to complete.

Our definition of the Fourier transform, when examined in another light, looks like an intertwining of representations of the affine group on $L^2(\mathbb{R})$. The affine group can be considered to be the set of first degree polynomials:

$$\{ax + b : a, b \in \mathbb{R} \text{ and } a \neq 0\},$$

with a group operation of composition. This may well allow group representation results to be exploited in this definition.

Chapter 3 also offers several questions in its conclusion. Can we use this type of construction for other transforms? Can we establish a convolution result using a different MRA, or a structure linked with MRA?

Finally Chapter 4 shows how little is known about dilation equations — a subject which looks remarkably similar to the familiar turf of difference equations and differential equations. Even cataloging all the solutions of the most simple dilation equation:

$$f(x) = f(2x) + f(2x-1),$$

does not seem to have been achieved! A more detailed study of dilation equations on a variety of spaces might level the playing field a little. This study might even produce new wavelets which are suitable for new situations.

Also dilation equations are only the first step in a chain where sums are replaced with integrals, functions replaced with distributions, scalars are replaced with matrices and single scales replaced with multiple scales. All in all there are not just many questions to be answered, but also many questions still to be thought of.

Bibliography

[1] Murtaza Ali and Ahmed H. Tewfik, *Multiscale difference equation signal models: Part 1 — theory*, IEEE Transactions on Signal Processing **43** (1995), no. 10, 2332–2345.

[2] J. M. Ash (ed.), *Studies in harmonic analysis*, Mathematical Association of America, 1976.

[3] Ingrid Daubechies, *Orthonormal bases of compactly supported wavelets*, Communications on Pure and Applied Mathematics **41** (1988), 909–996.

[4] Ingrid Daubechies and Jeffrey C. Lagarias, *Two-scale difference equations. 1: Existence and global regularity of solutions*, SIAM Journal of Mathematical Analysis **22** (1991), no. 5, 1388–1410.

[5] ———, *Two-scale difference equations. 2: Local regularity, infinite products of matricies and fractals*, SIAM Journal of Mathematical Analysis **23** (1992), no. 4, 1031–1079.

[6] Mathsoft Incorporated, *Wavelet resources*, `http://www.mathsoft.com/wavelets`.

[7] Ian Johnston, *Measured tones*, Institute of Physics, Bristol, 1993.

[8] Gerald Kaiser, *A friendly guide to wavelets*, Birkhäuser, Boston, 1994.

[9] Tom Koornwinder (ed.), *Wavelets: An elementry treatment of theory and applications*, World Scientific Publishing, Singapore, 1993.

[10] Erwin Kreysig, *Advanced engineering mathematics*, sixth ed., John Wiley & Sons, New York, 1988.

[11] W. Christopher Lang, *Orthogonal wavelets on the cantor dyadic group*, SIAM Journal of Mathematical Analysis **27** (1996), no. 1, 305–312.

[12] Yves Meyer, *Wavelets and operators*, Cambridge University Press, Cambridge, 1992.

[13] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery, *Numerical recipies in C*, second ed., Cambridge University Press, Cambridge, 1992.

[14] Walter Rudin, *Real and complex analysis*, second ed., McGraw-Hill, New York, 1974.

[15] ______, *Functional analysis*, second ed., McGraw-Hill, New York, 1991.

[16] Elias M. Stein and Guido Weiss, *Introduction to fourier analysis on euclidean spaces*, Princeton University Press, Princeton, 1971.

[17] Gilbert Strang, *Wavelet transforms versus fourier transforms*, Bulletin of the AMS **28** (1993), no. 2, 288–305.

[18] Gilbert Strang and Truong Nguyen, *Wavelets and filter banks*, Wellesley-Cambridge Press, Massachusetts, 1996.

[19] A. H. Stroud, *Numerical quadrature and solutions of ordinary differential equations*, Springer, 1974.

[20] Brani Vidaković and Peter Müller, *Wavelets for kids*, `ftp://ftp.isds.duke.edu/pub/Users/brani/wav4kidsA.ps.Z`.

[21] Lars F. Villemoes, *Energy moments in time and frequency for two-scale difference equation solutions and wavelets*, SIAM Journal of Mathematical Analysis **23** (1992), no. 6, 1519–1543.

[22] Alan Watt and Mark Watt, *Advanced animation and rendering techniques*, Addison-Wesley, Wokingham, 1992.