CHAPTER 1

Introduction and Overview

In this section, I will try to address the following questions:

• What are the ideas behind transforms in general?
• What is the idea behind the wavelet transform?

1.1. Transforms

A transform is a mapping which takes a function (or a sequence of numbers) and maps it into another function (or sequence of numbers). Reasons for taking a transform include

• The values of the transform may give us some information on the original function, such as smoothness, rate of decay, periodicity, etc.
• The transform of an equation may be easier to solve than the original equation.
• The transform of a vector may require less storage.
• We may want to apply some operation (such as smoothing) to the original function, and find that it is easier to do on the transform side.

Most useful transforms are linear, one-to-one, and invertible. This means

• linear: We can pull out constants and transform the terms of a sum separately:

\[ T(\alpha f + \beta g) = \alpha (Tf) + \beta (Tg). \]

Here we assume that \( f, g \) are functions (or sequences), \( \alpha, \beta \) are numbers, and \( Tf \) is the transform of \( f \).

• one-to-one: If \( f, g \) are different functions (sequences), then their transforms \( Tf, Tg \) are different.

• invertible: There is an inverse transform \( T^{-1} \) which recovers \( f \) from \( Tf \).

I will use the word continuous transform to denote one that maps functions to functions. The word “continuous” usually means something else in mathematics, but that should not cause any confusion in this context. A discrete transform maps (finite or infinite) sequences to sequences. There are also some semidiscrete transforms that relate functions with sequences.

Most continuous linear transforms have the form

\[ (Tf)(\xi) = \int k(\xi, x)f(x) \, dx, \]

where the function \( k \) is called the kernel. If the kernel depends only on \( (\xi - x) \), not on \( \xi \) and \( x \) separately, this is called a convolution (or filter, for the engineers):

\[ (Tf)(\xi) = \int k(\xi - x)f(x) \, dx. \]

For discrete transforms, the counterparts are

\[ (Tf)_i = \sum_j k_{ij}f_j \]

in general, or

\[ (Tf)_i = \sum_j k_{i-j}f_j \]

in the case of a discrete convolution.

Example: The Fourier Transform (continuous) is defined by

\[ \hat{f}(\xi) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-ix\xi}f(x) \, dx. \]
The \textit{Laplace Transform} (continuous) is defined by
\[ \mathcal{L}f(s) = \int_0^\infty e^{-st}f(t) \, dt. \]

The \textit{Discrete Fourier Transform} (discrete) of a vector of length \( N \) is defined by
\[ \hat{f}_k = \sum_{j=0}^{N-1} e^{-i \frac{2\pi}{N} k j} f_j, \quad k = 0, \ldots, N - 1. \]

All these transforms are linear, one-to-one, and have an inverse. None of them is a convolution. \( \square \)

One way to interpret the values of a transform is to consider them as inner products of the original function with various “test functions”.

Recall that if \( f, g \) are finite-dimensional real vectors, their inner product satisfies
\[ \langle f, g \rangle = \|f\|_2 \cdot \|g\|_2 \cdot \cos \theta, \]
where \( \theta \) is the angle between \( f \) and \( g \). (The inner product is defined in appendix B). Thus, \( \langle f, g \rangle \) is large if \( f, g \) are almost parallel, and small if \( f, g \) are close to perpendicular. If \( g \) has length 1, \( \langle f, g \rangle \) can be interpreted as the size of the component of \( f \) in direction \( g \).

Now consider a discrete transform
\[ (Tf)_i = \sum_j k_{ij} f_j \]
For fixed \( i \), \( k_i = \{k_{ij}\}_j \) is a vector, and
\[ (Tf)_i = \langle f, \tilde{k}_i \rangle, \]
so the \( i \)th component of the transformed vector \( Tf \) is the component of \( f \) in direction \( \tilde{k}_i \). Taking the transform of \( f \) is equivalent to “testing” \( f \) with many different vectors, to see how much \( f \) “looks like” the test vectors.

The same idea, with minor modifications, applies to continuous transforms as well.

\textbf{Example: The Discrete Fourier Transform.} The test vectors are complex, but their real and imaginary parts are discretized versions of sin and cos functions of various frequencies:

Thus, a coefficient in the DFT of \( f \) roughly corresponds to the frequency component of \( f \) at the corresponding frequency. \( \square \)
1.2. The Wavelet Transform

This section illustrates one of the reasons for introducing the Continuous Wavelet Transform.

Suppose you want to analyze a time-varying signal. The Fourier Transform tells you what frequencies are present in the signal.

**Example:** The data represents the number of sunspots recorded in the years 1700 through 1987 (taken from [KMN89]). The Fourier Transform is complex. The plot shows the real and imaginary parts of the FT, and the power spectrum (absolute value of FT). Choosing a frequency axis from 0 to 1 means that a frequency $\xi$ corresponds to a period of $1/\xi$ years. (The values in the second half of the FT represent negative frequencies; for real data, this is just a complex conjugate mirror image of the values in the first half).

It is well known that sunspots have a cycle of about 11.2 years, which shows up as a distinct spike in the FT (best seen in the power spectrum) around $1/11.2$.

![Sunspot Data](image)

The Fourier Transform works well for periodic signals. However, it does not yield any information on time localization of frequencies. Signals whose frequency content varies with time cannot be analyzed properly this way.

**Example:** Consider the Fourier Transforms of two signals, both composed of two pure sin functions of different frequencies. On the left, each frequency is present by itself for half of the interval. On the right, the two frequencies are both present the entire time. The Fourier transforms show similar peaks.
This problem is of course related to the fact that all the “test functions” employed in the Fourier Transform cover the entire interval. The Wavelet Transform, in contrast, employs localized test functions.

Taking a Wavelet Transform is equivalent to “testing” the original function with compressed and shifted versions of a single wave form, the mother wavelet. Some sample wavelets are shown in the picture below. Notice how the more compressed wavelets have smaller shifts, so it takes more of them to cover a given interval.
As an example, consider the wavelet transform of the two signals in the last example. I used the so-called Haar wavelets for this, whose mother wavelet looks like this:

A more detailed explanation of what the wavelet transform means has to wait until later, but compare regions 1 and 2 in the two transform pictures; they correspond to the main frequencies present. In the
Wavelet Transform, you can tell whether something is happening in the left or right half only, or in the whole region.

1.3. Suggested Reading

Chapter 1 in Daubechies [Dau92]. Chapter 1 in Chui [Chu92]. The survey article by Jawerth and Sweldens [JS].

1.4. Bibliography


