Notes on

Generalized Functions

and

Fourier Transforms

A compendium

by
J Bergh

Göteborg 1994
(translated 2002)

Preface

This compendium is meant as a terse replacement for, *i.a.*, mathematically less successful parts (Chapter 5, for example) of the book Bracewell R., *The Fourier Transform and Its Applications,* McGraw-Hill, 2000.

In part, we transcend this book: the results named after Paley-Wiener and Bochner-Schwartz are included together with some theory of generalized functions: *tempered distributions.*

We mention two books for further study of distributions.

Folland G. B., *Fourier Analysis and its Applications,* Wadsworth & Brooks, 1992

Hörmander L., *The Analysis of Linear Partial Differential Operators* part I, second edition, Springer-Verlag, 1990

The first book is fairly easy to read. The second contains a more comprehensive account. Both are available in the library at the Department, as well as many more on the subject.

Our disposition is as follows. We start with the class $\mathcal{S}$ of infinitely differentiable fuctions with rapid decay at infinity. The Fourier transform is an isomorphism on $\mathcal{S}$. Then we discuss the dual of $\mathcal{S}$, $\mathcal{S}'$, which are generalized functions called tempered distributions. As applications of the calculus within this framework, we give relatively straight-forward proofs of Poisson's Summation Formula, the Sampling Theorem, convergence of Fourier series, and the Central Limit Theorem. We characterize also the functions which have no frequency content above a fixed value (Paley-Wiener), and the connection between autocorrelation functions and probability measures (Bochner-Schwartz). We briefly discuss the Radon transform (used in computer tomography and several other contexts), antennas, and thin lenses. This is followed by some issues pertaining to the transition between a continuous variable and its finite discrete counterpart which can be handled by a computer.

Finally we sketch the idea behind the *wavelet transform* – a variant of (windowed) Fourier transform which is becoming widely used in diverse applications (storing fingerprints for example).

Mainly, we will use the notation of Bracewell. However, all results have versions in more than one dimension, and the proofs in higher dimensions do not, in general, require any additional ideas.

In many places in the text, there are exhortations like 'verify!'. This means that some, mostly minor and technical, details have been left out. The purpose with these gaps is above all to make the ideas stand out more clearly; most of the gaps will be discussed during the course.

Contents

1 Generalized Functions

We will expand the concept *function*, and then conceive a function f as the values of

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

as φ runs through a class of *test functions*. Note that this differs from the usual way of looking at a function as, *e.g.,* a graph.

This section starts with the test functions, which make up the class $\mathcal{S}$.

We then treat the *tempered distributions*, $\mathcal{S}'$, an expansion of the function concept, which simplifies calculations with, *i.a.*, sampled signals.[1]

We end this section with some applications. Among other, we give simple proofs of the Sampling Theorem, the Poisson Summation Formula. We also discuss the connection between Fourier transforms and Fourier series.

When do distributions come more specifically into play?

Mathematical techniques are used to facilitate manipulation of mathematical models, which fit 'reality' more or less, as the case may be. From signal processing we take two examples where distributions help. (Many applications can be found in the theory of partial differential equations; cf. Example 2 in 2.2 below.)

1. We seek a signal of which the 'moving averages' are known (or perhaps the moving averages of those). Put differently, we seek the solution f to the equation $1_{(-1/2,1/2)} * f = g \quad (1_{(-1/2,1/2)} * 1_{(-1/2,1/2)} * f = h)$.
2. Consider a low-pass filtered signal: no frequency content above the 'Nyquist frequency'. Could this signal have maxima as close to each other as we wish, or is there a minimal distance determined by the Nyquist frequency?

1.1 The Function Class $\mathcal{S}$

Definition 1.1.1 *The* ***class*** $\boldsymbol{\mathcal{S}}$ *consists of complex-valued functions f of (one) real variable, which are infinitely differentiable and satisfy*

$$\sup_x |x^\alpha D^\beta f(x)| < \infty$$

for each choice of non-negative integers α, β.

In other words: *the function and all its derivatives decay faster than any power function at infinity.*

[1] Distributions were developed as an aid to the study of linear partial differential equations and their solutions.

EXAMPLE 1 Let $f(x) = e^{-ax^2}$ with $a > 0$. Then $f \in \mathcal{S}$ holds, which may be verified directly from the definition.

EXAMPLE 2 Verify also that $f \in \mathcal{S}$ when f is given by (draw a picture!)

$$f(x) = \begin{cases} 0 & (x \leq a) \\ e^{-\frac{1}{(x-a)^2} - \frac{1}{(x-b)^2}} & (a < x < b) \\ 0 & (b \leq x) \end{cases}$$

EXAMPLE 3 Let f be the function in the previous example with $a = -b$ and normalized to have integral 1. Let g be an integrable function.[2] Define the convolution ($\varepsilon > 0$)

$$g_\varepsilon(x) = \int_{-\infty}^{\infty} \frac{1}{\varepsilon} f(\frac{x-y}{\varepsilon}) g(y)\, dy$$

Then g_ε is infinitely differentiable, and converges to g almost everywhere (verify!).

Definition 1.1.2 *The **Fourier transform** of a function f in $\mathcal{S}$ we denote by $\hat{f}$ or $\mathcal{F}f$. It is given by*

$$\hat{f}(s) = \int_{-\infty}^{\infty} e^{-2\pi i s x} f(x)\, dx$$

The properties which make the class $\mathcal{S}$ suitable for Fourier transforms are contained in the next lemma.

Lemma 1.1.1 *Let $f \in \mathcal{S}$. Then holds for non-negative integers α, β*

1. $g(x) = x^\alpha D^\beta f(x) \Rightarrow g \in \mathcal{S}$

   "$\mathcal{S}$ is closed under differentiation and under multiplication by polynomials"

2. $\hat{f} \in \mathcal{S}$

   "$\mathcal{S}$ is closed under Fourier transformation"

Proof: To prove the last property, it suffices to observe the following two equalities (verify this!). Firstly, differentiation under the integral sign gives

$$D\hat{f}(s) = \int_{-\infty}^{\infty} e^{-2\pi i s x}(-2\pi i x) f(x)\, dx$$

[2] Lebesgue integrable for example. If this concept is unfamiliar, read 'continuous and Riemann integrable' and 'everywhere'.

The operation is allowed, since the resulting integral is absolutely and uniformly convergent. Thus we have $\sup_s |D^\beta \hat{f}(s)| < \infty$ for each $\beta \geq 0$. Secondly, we get by a partial integration

$$2\pi i s \hat{f}(s) = \int_{-\infty}^{\infty} e^{-2\pi i s x} Df(x)\, dx$$

which yields $\sup_s |s^\alpha \hat{f}(s)| < \infty$ for every $\alpha \geq 0$.

Verify the first property!

□

EXAMPLE 4 Put $f(x) = e^{-\pi x^2}$. Then $\hat{f} = f$. We verify this:

$$\begin{aligned} D\hat{f}(s) &= \int_{-\infty}^{\infty} e^{-2\pi i s x}(-2\pi i x)e^{-\pi x^2}\, dx \\ &= i \int_{-\infty}^{\infty} e^{-2\pi i s x} Df(x)\, dx \\ &= -2\pi s \hat{f}(s) \end{aligned}$$

which implies $\hat{f}(s) = e^{-\pi s^2}$ since $\hat{f}(0) = \int f(x)\, dx = 1$.

We will now show the key result: a function in $\mathcal{S}$ can be retrieved from its Fourier transform.

Theorem 1.1 (Fourier's Inversion Formula) *Let $f \in \mathcal{S}$. Then holds*

$$f(x) = \int_{-\infty}^{\infty} e^{2\pi i s x} \hat{f}(s)\, ds$$

REMARK 1

1. The formula is valid in the distribution sense under weaker assumptions, *e.g.*, for $f \in L^2$ *(square integrable)*. Cf. Theorem 1.3 below.

2. The formula may be written $\mathcal{F}^2 f(x) = f(-x)$, which means that four successive Fourier transformations produce the original function.

   We write $\check{f}(x) := f(-x)$; that is, $\mathcal{F}^2 f(x) = \check{f}(x)$.

3. Lemma 1.1.1 and Theorem 1.1 show that the Fourier transform is an isomorphism on $\mathcal{S}$. (This includes the topology; see the proof of Proposition 1.2.2. below.)

Proof: It suffices to consider the case $x = 0$ (verify!).

Suppose first that $f(0) = 0$. We will then show that $\int \hat{f}(s)\, ds = 0$. Put $g(x) = f(x)/x$ and $g \in \mathcal{S}$ follows (verify!). Furthermore $-2\pi i \hat{f}(s) = D\hat{g}(s)$ and thus

$$-2\pi i \int \hat{f}(s)\, ds = \int D\hat{g}(s)\, ds = 0$$

which proves the theorem in the case $f(0) = 0$.

If $f(0) \neq 0$, write

$$f(x) = f(x) - f(0)e^{-\pi x^2} + f(0)e^{-\pi x^2}$$

Taking Fourier transform, integrating, and using what we just showed, we get

$$\int \hat{f}(s)\, ds = f(0)$$

because the inversion formula holds for $e^{-\pi x^2}$ according to Example 4 above. The proof is done. $\square$

REMARK 2 The proof displays a technique which is frequently used: You split the proof into two steps. First you show the statement for $f \in \mathcal{S}$ with $f(0) = 0$, that is, for a subspace of $\mathcal{S}$. Then a general function is split into a sum of two terms: one term in the subspace and the behaviour of the other term (not in the subspace) is also known. This then gives the general statement in a linear setting.

1.2 The Class $\mathcal{S}'$

We will now describe a generalization of the *function* concept, named *tempered distribution*,[3] which allows a unified treatment of ideas like *point mass, point charge, impulse, shot signal, dipole moment etc.*

Test functions will be denoted by Greek letters, for example φ, in what follows, while distributions will be written with capitals in italics, *e.g.*, T. The complex numbers we denote $\mathbf{C}$ (in boldface). Furthermore we use the notation (φ_1, φ_2 in $\mathcal{S}$)

$$< \varphi_1, \varphi_2 > := \int_{-\infty}^{\infty} \varphi_1(x)\varphi_2(x)\, dx$$

Note that $< \cdot, \cdot >$ is a scalar product only for real-valued functions.

Definition 1.2.1 *A linear mapping*

$$T: \ \mathcal{S} \ni \varphi \longrightarrow T(\varphi) \in \mathbf{C}$$

[3] The adjective 'tempered' denotes here 'tempered (moderate) growth'. See Example 8 below.

is called a **tempered distribution** *if, for any sequence of test functions $\varphi_n \in \mathcal{S}$ with the property*

$$\lim_{n\to\infty} \sup_x |x^\alpha D^\beta \varphi_n(x)| = 0$$

for every choice of non-negative integers α, β, holds that

$$\lim_{n\to\infty} T(\varphi_n) = 0$$

The class of tempered distributions is written $\mathcal{S}'$; and we often write $T(\varphi) =: < T, \varphi >$; in Example 5 below the reason for this will become apparent. A sequence of test functions with the property above is said to *converge to* 0 *in* $\mathcal{S}$.

An example of such a sequence is $\varphi_n(x) = e^{-x^2}/n$ with $n = 1, 2, \dots$. Verify that $\varphi_n \to 0$ i $\mathcal{S}$.

Verify also that *a tempered distribution is determined by its values on real-valued test functions only.*

EXAMPLE 5 Let $|f(x)|/(1+x^2)^\alpha$ be integrable for some α and put

$$T:\ \mathcal{S} \ni \varphi \longrightarrow < f, \varphi > = \int_{-\infty}^{\infty} f(x)\varphi(x)\, dx \ \in \mathbf{C}$$

We verify that the mapping T is a tempered distribution. The linearity is obvious.

Take a sequence $\varphi_n \to 0$ i $\mathcal{S}$. Then we get

$$|\int_{-\infty}^{\infty} f(x)\varphi_n(x)\, dx| \le \int_{-\infty}^{\infty} |f(x)/(1+x^2)^\alpha|\, dx \sup_x |(1+x^2)^\alpha \varphi_n(x)| \to 0$$

which finishes the verification. Note that f could be a polynomial here for example.

The distribution T in Example 5 is identified with the function f. This identification has often proved difficult to get used to.

The function f is thus here not conceived of as the values $f(x)$ as x varies, but as the values $< f, \varphi >$ as φ varies over $\mathcal{S}$. We write $T = f$ in $\mathcal{S}'$.

Verify that, if f and g both are continuous functions and $f = g$ in $\mathcal{S}'$ ('equal as distributions'), that is $< f, \varphi > = < g, \varphi >$ for all $\varphi \in \mathcal{S}$, then $f(x) = g(x)$ holds for all x.

In the next example we treat a frequently used tempered distribution, the *δ-distribution.*[4]

EXAMPLE 6 Consider the mapping

$$T:\ \mathcal{S} \ni \varphi \longrightarrow \varphi(0) \in \mathbf{C}$$

Verify that T is a tempered distribution as was done in Example 5.

[4] The notation derives from the name P A M Dirac.

Remark 3 We counsel against the usage of the (abusive) notation like 'δ-function', '$\delta(x)$', '$\int \delta(x)\varphi(x)\,dx$', and so on. These suggest the erroneous conception that the δ-distribution has point values like functions, and introduce unnecessary possibilities for misunderstanding. See also the Structure Theorem 1.2 below.

Example 7 Put $f_n(x) = n^{1/2}e^{-n\pi x^2}$ with $n = 1, 2, \ldots$. Verify that, $f_n \in \mathcal{S}'$ and

$$f_n(\varphi) \to \varphi(0) = \delta(\varphi)$$

This may be written $f_n \to \delta$ in $\mathcal{S}'$. We can thus say that f_n approximate δ.

Example 8 Continuous functions which are not *tempered* distributions grow too fast at infinity: consider e^x, for example. Take $\varphi(x) = e^{-(1+x^2)^{1/2}}$. Obviously, $\varphi \in \mathcal{S}$ but $< e^{(\cdot)}, \varphi >$ diverges, and so $e^{(\cdot)}$ does not belong to $\mathcal{S}'$.[5]

Example 9 In signal processing, the most widely used tempered distribution is the *pulse train* $\sum_n \delta_n$, where $\delta_n(\varphi) = \varphi(n)$, that is, δ translated to the integer n.[6] Se Remark 7 below.

Verify that the pulse train, which is the mapping

$$T:\ \mathcal{S} \ni \varphi \longrightarrow \sum_n \varphi(n) \in \mathbf{C}$$

is a tempered distribution. (Note that $< \sum_n \delta_n, \varphi > := \sum_n \varphi(n)$.)

The operations *differentiation, multiplication by a function, translation*, and others, must be defined in such a way that they coincide with the usual ones when the distribution is a function. For example, if $f \in \mathcal{S}$ and $\varphi \in \mathcal{S}$ then

$$< Df, \varphi > = \int Df(x)\varphi(x)\,dx = -\int f(x)D\varphi(x)\,dx = -< f, D\varphi >$$

The other definitions below are motivated similarly.
Translation is denoted $f_\tau(x) := f(x-\tau)$, τ real. Multiplication by a function g is defined under the weakest assumption on g for which the implication $\varphi \in \mathcal{S} \Rightarrow g\varphi \in \mathcal{S}$ holds true.

Definition 1.2.2 *Let $T \in \mathcal{S}'$, $\varphi \in \mathcal{S}$, and let g be an infinitely differentiable function with the property that, given the integer $\beta \geq 0$ there is an integer α such that $\sup_x (1+|x|)^\alpha |D^\beta g(x)| < \infty$.*

$$\begin{aligned} < DT, \varphi > &:= -< T, D\varphi > \\ < gT, \varphi > &:= < T, g\varphi > \\ < T_\tau, \varphi > &:= < T, \varphi_{-\tau} > \end{aligned}$$

[5] Notation like $e^{(\cdot)}$, *i.e.*, with the variable suppressed, is used to diminish the risk for misunderstandings, in particular when we later discuss multiplication of a tempered distribution by a function.

[6] Bracewell's notation with the Russian letter III, *'shah'*, is awkward especially in connection with changes of variable, and we will not use it.

Polynomials are thus allowed as g in the definition, but not e^x, for example. The objects defined should be tempered distributions, which is the content of the next proposition.

Proposition 1.2.1 *For $T \in \mathcal{S}'$ and with g as in Definition 1.2.2 hold $DT \in \mathcal{S}'$, $gT \in \mathcal{S}'$ and $T_\tau \in \mathcal{S}'$.*

Proof: We prove the statement about gT, and leave the other two as an exercise.

Linearity is obvious. Take now a sequence $\varphi_n \to 0$ in $\mathcal{S}$,

$$< gT, \varphi_n > = < T, g\varphi_n >$$

But, given α, β, we have

$$\sup_x (1+|x|)^\alpha |D^\beta (g(x)\varphi_n(x))| \leq C \sum_{\nu \leq \beta} \sup_x (1+|x|)^{\alpha_\nu} |D^\nu \varphi_n(x)|$$

which implies $g\varphi_n \to 0$ in $\mathcal{S}$, and $gT \in \mathcal{S}'$ follows. □

Proposition 1.2.1 can be rephrased: *The class $\mathcal{S}'$ is closed under differentiation, under multiplication with smooth functions which have tempered growth at infinity, and under translation.*

Note that polynomials satisfy the conditions for g in Proposition 1.2.1.

We will now indicate a representation of a general tempered distribution in terms of (distribution) derivatives of continuous functions. The example following the theorem provides an illustration.

Theorem 1.2 (The Structure Theorem) *Let $T \in \mathcal{S}'$. Then continuous functions f_j, $j = 1, 2, \ldots$, and non-negative integers β_j exist, such that (in $\mathcal{S}'$)*

$$T = \sum_j D^{\beta_j} f_j$$

Proof: See Appendix. □

EXAMPLE 10 For $f(x) = x_+$, the *ramp function*, and the Heaviside function *(step function)* H, $D^2 f = DH = \delta$ in $\mathcal{S}'$ holds. We verify this. ($\varphi \in \mathcal{S}$)

$$\begin{aligned} < D^2 f, \varphi > &= < f, D^2\varphi > \\ &= \int_0^\infty x D^2\varphi(x)\, dx \\ &= -\int_0^\infty D\varphi(x)\, dx \; (= < DH, \varphi >) \\ &= \varphi(0) = < \delta, \varphi > \end{aligned}$$

EXAMPLE 11 Let f be the function $f(x) = -x^{-3/2}H(x)/2$, where H again is the Heaviside function. Define

$$< T, \varphi > := \lim_{\epsilon \to 0^+} \{ -\frac{1}{2} \int_\epsilon^\infty x^{-3/2} \varphi(x)\, dx + \epsilon^{-1/2} \varphi(0) \}$$

T is called the *finite part* of f.

Verify that the continuous function $g = 2(\cdot)^{1/2} H \in \mathcal{S}'$, that $Dg = (\cdot)^{-1/2} H$, and that $D^2 g = T$ in $\mathcal{S}'$! This motivates the definition of T. The added terms in the finite part of a function come from a series expansion of the test function $\varphi(x)$.

We are now ready to define the Fourier transform of a tempered distribution. Again, the definition is motivated by a formula for functions, the *Plancherel Formula* ($f \in \mathcal{S},\ \varphi \in \mathcal{S}$)

$$\int \mathcal{F} f(x) \varphi(x)\, dx = \int f(x) \mathcal{F} \varphi(x)\, dx$$

The formula is verified directly by changing the order of integration. Note the special case $\varphi = (\hat{f})^*$, which is called *Parseval's Formula:*[7]

$$\int |\hat{f}(s)|^2\, ds = \int |f(x)|^2\, dx$$

Definition 1.2.3 *Take $T \in \mathcal{S}'$. The* ***Fourier transform*** *of T, $\hat{T} = \mathcal{F}T$, is given by*

$$< \hat{T}, \varphi > := < T, \hat{\varphi} >$$

Proposition 1.2.2 *$T \in \mathcal{S}'$ implies $\hat{T} \in \mathcal{S}'$.*

Proof: Linearity is obvious. Take test functions $\varphi_n \to 0$ in $\mathcal{S}$ and

$$< \hat{T}, \varphi_n > = < T, \hat{\varphi_n} >$$

If we have $\hat{\varphi_n} \to 0$ in $\mathcal{S}$, we are done. We have both

$$\sup_s |D\hat{\varphi_n}(s)| \le C \int |x \varphi_n(x)|\, dx$$

and

$$\sup_s |s\hat{\varphi_n}(s)| \le C \int |D\varphi_n(x)|\, dx$$

But for the first integrand holds for example

$$|x\varphi_n(x)| \le (1 + x^2)^{-1} \sup_x (1 + x^2)^2 |\varphi_n(x)|$$

A similar argument for the last integrand gives $\hat{\varphi_n} \to 0$ in $\mathcal{S}$, and the proof is complete. (Verify the last estimate!) □

[7] It is called Rayleigh's Theorem in Bracewell's book, but this naming is uncommon.

EXAMPLE 12 It is immediately seen that ($\beta \geq 0$)

$$\mathcal{F}D^\beta\delta = (2\pi i(\cdot))^\beta$$

$$\mathcal{F}(-2\pi i(\cdot))^\beta = D^\beta\delta$$

We verify the first statement for $\beta = 0$:

$$< \mathcal{F}\delta, \varphi > = < \delta, \mathcal{F}\varphi > = \mathcal{F}\varphi(0) = \int \varphi(x)\, dx$$

The rest is left as an exercise.

The central result is Fourier's Inversion Formula, where $\check{T}$ is defined by

$$< \check{T}, \varphi > := < T, \check{\varphi} >$$

Verify that $\check{T} \in \mathcal{S}'$.

Theorem 1.3 (Fourier's Inversion Formula) *Let $T \in \mathcal{S}'$. Then $\mathcal{F}\mathcal{F}T = \check{T}$ holds.*

Proof: For $\varphi \in \mathcal{S}$ we have the same formula proved in Theorem 1.1, and so

$$\begin{aligned} < \mathcal{F}^2T, \varphi > &= < \mathcal{F}T, \mathcal{F}\varphi > \\ &= < T, \mathcal{F}\mathcal{F}\varphi > \\ &= < T, \check{\varphi} > \\ &= < \check{T}, \varphi > \end{aligned}$$

The proof is complete. $\square$

The Fourier transform effects of differentiation, translation, and multiplication by certain functions in Proposition 1.2.1, is next. Recall the corresponding rules for functions in $\mathcal{S}$!

Proposition 1.2.3 *The Fourier transform is linear. Let $T \in \mathcal{S}'$, and assume that S is an infinitely differentiable function, such that, given $\beta \geq 0$, α exists, and $\sup_x(1+|x|)^\alpha|D^\beta S(x)| < \infty$.*

The following equalities hold.

$$\begin{array}{ll} \mathcal{F}(DT) = 2\pi i(\cdot)\mathcal{F}T & \mathcal{F}(-2\pi i(\cdot)T) = D\mathcal{F}T \\ \mathcal{F}(ST) = \mathcal{F}S * \mathcal{F}T & \mathcal{F}((\mathcal{F}S) * T) = \check{S}\mathcal{F}T \\ \mathcal{F}(T_\tau) = e^{-2\pi i\tau(\cdot)}\mathcal{F}T & \mathcal{F}(e^{2\pi i\tau(\cdot)}T) = (\mathcal{F}T)_\tau \end{array}$$

where the convolutions are defined in the proof below. Furthermore,

$$D(\mathcal{F}S * T) = D\mathcal{F}S * T = \mathcal{F}S * DT$$

Proof: The linearity together with the first and the third equalities are directly verified from the definition. Likewise, $S \in \mathcal{S}'$. (Perform the verifications!)

We will now define convolution of a general tempered distribution $\mathcal{F}T \in \mathcal{S}'$ with a tempered distribution $\mathcal{F}S \in \mathcal{S}'$ with the property stated. The product ST is in $\mathcal{S}'$, and we define $\mathcal{F}S * \mathcal{F}T$ by $\mathcal{F}(ST) =: \mathcal{F}S * \mathcal{F}T$; the Fourier transform determines the tempered distribution completely. The second equality on that line follows from the definition (the first equality). The commutativity of convolution, translation, and differentiation (the last line of the statement) follows after Fourier transformation of what we just have shown, and since multiplication by certain functions is associative in $\mathcal{S}'$ when it is permitted. $\square$

Now we present a result, which is basic in most applications. It provides the answer to the question *What is the result after a division by x ?*

Lemma 1.2.1 [8] *Let $T \in \mathcal{S}'$, and assume that $(\cdot)T = 0$. Then there is a complex number a such that*

$$T = a\delta$$

Performing a Fourier transform in the lemma produces:

Consequence 1.2.1 *Let $T \in \mathcal{S}'$, and assume that $DT = 0$. Then there is a complex number a such that $T = a$.*

Proof of Lemma 1.2.1: Assume that $\psi \in \mathcal{S}$ with $\psi(0) = 0$. For $\varphi(x) = \psi(x)/x$ then $\varphi \in \mathcal{S}$ holds (verify!), and

$$T(\psi) = T((\cdot)\varphi) = (\cdot)T(\varphi) = 0$$

Take now $\varphi \in \mathcal{S}$ arbitrarily, fix $\varphi_1 \in \mathcal{S}$ with $\varphi_1(0) = 1$, and write

$$\varphi(x) = \varphi(x) - \varphi(0)\varphi_1(x) + \varphi(0)\varphi_1(x)$$

this yields $(\varphi(0) - \varphi(0)\varphi_1(0) = 0)$

$$T(\varphi) = \varphi(0)T(\varphi_1) = T(\varphi_1)\delta(\varphi)$$

which proves the lemma with $a = T(\varphi_1)$. $\square$

[8] This result is not immediately generalized to dimension 2 and higher.

EXAMPLE 13 Let H be the Heaviside function. Then

$$\mathcal{F}H = \frac{1}{2\pi i(\cdot)} + \frac{1}{2}\delta$$

where

$$< \frac{1}{2\pi i(\cdot)}, \varphi >= \lim_{\epsilon \to 0+} \int_{|x|>\epsilon} \frac{1}{2\pi i x}\varphi(x)\, dx$$

(verify that $(2\pi i(\cdot))^{-1} \in \mathcal{S}'$!).

We show that

$$\mathcal{F}H_1 = (2\pi i(\cdot))^{-1}$$

with $H_1 = H - 1/2$.[9] Recall that

$$DH_1 = DH = \delta$$

which implies

$$2\pi i(\cdot)\mathcal{F}H_1 = 1$$

The lemma gives

$$\mathcal{F}H_1 = (2\pi i(\cdot))^{-1} + a\delta$$

Note that $\mathcal{F}H_1$ and $(2\pi i(\cdot))^{-1}$ both are odd $((\mathcal{F}H_1)\check{} = -\mathcal{F}H_1)$ while δ is even (verify!). Thus we infer $a = 0$, and the proof is done.

REMARK 4 The theory needed to answer the two questions posed (on page 1) as specific examples when distributions are useful has now been described.

1.3 Some Applications

We will apply the theory developed in the foregoing to show key mathematical results used for signal processing.

Theorem 1.4 (Poisson's Summation Formula) *Let $\varphi \in \mathcal{S}$. Then*

$$\sum_{k=-\infty}^{\infty} \hat{\varphi}(s+k) = \sum_{k=-\infty}^{\infty} \varphi(k) e^{-2\pi i k s}$$

or, equivalently,

$$\mathcal{F}\{\sum_{k=-\infty}^{\infty} \delta_k\} = \sum_{k=-\infty}^{\infty} \delta_k$$

[9] $2H_1 =: \text{sign}$, where sign is called the *sign function*.

We give first a proof in dimension 1. Using Theorem 1.5 (which has other elementary proofs using partial sums) below, we return (after Proposition 1.3.2) and give a second proof, that works unaltered in higher dimensions.

Proof: We have

$$\begin{cases} (\sum_k \delta_k)_1 = \sum_k \delta_k & \text{(period 1)} \\ e^{2\pi i(\cdot)} \sum_k \delta_k = \sum_k \delta_k & (e^{2\pi i(\cdot)}\delta_k = \delta_k) \end{cases}$$

A Fourier transformation gives

$$\begin{cases} e^{2\pi i(\cdot)} \mathcal{F} \sum_k \delta_k = \mathcal{F} \sum_k \delta_k \\ ((\mathcal{F} \sum_k \delta_k)_1 =) \, (\mathcal{F} \sum_k \delta_k)_{-1} = \mathcal{F} \sum_k \delta_k \end{cases}$$

Lemma 1.2.1 now gives constants a_k, which all must be equal by the periodicity (translation property), so that

$$\mathcal{F} \sum_k \delta_k = a \sum_k \delta_k$$

This equality applied to the test function $e^{-\pi x^2}$, which is its own Fourier transform, gives $a = 1$. This finishes the proof.

$\square$

EXAMPLE 14 Let $\varphi \in \mathcal{S}$ with $\mathcal{F}\varphi(s) = 0, \; |s| \geq 1$. Then

$$\sum_n \varphi(n) = \int \varphi(x)\, dx$$

The integral can apparently be replaced by a rectangular approximation centered at the integers in this case!

Now a result concerning the convergence of Fourier series. A comparison between the proof we give and an elementary one might be profitable.

Theorem 1.5 *Suppose that the function f has period 1 and is twice continuously differentiable. Then*

$$f(x) = \sum_k c_k e^{2\pi i k x}$$

holds for all x, where the Fourier coefficients c_k are given by (k integer)

$$c_k = \int_0^1 e^{-2\pi i k x} f(x)\, dx$$

Proof: Two partial integrations give $c_k = \mathcal{O}(k^{-2}),\ |k| \to \infty$. The Fourier series

$$\sum_{k=-\infty}^{\infty} c_k e^{2\pi i k x}$$

then converges absolutely and uniformly. The sum is thus a continuous function with period 1 and so belongs to $\mathcal{S}'$.

It suffices to show equality in $\mathcal{S}'$, since both f and the Fourier series are separately continuous functions. Equality in $\mathcal{S}'$ is equivalent to

$$\hat{f} = \sum_k c_k \delta_k$$

Poisson's Summation Formula gives ($\varphi \in \mathcal{S}$)

$$\begin{aligned} < \hat{f}, \varphi > = \int f(x)\hat{\varphi}(x)\,dx &= \int_0^1 f(x) \sum_k \hat{\varphi}(x+k)\,dx \\ &= \int_0^1 f(x) \sum_k e^{-2\pi i k x} \varphi(k)\,dx \\ &= \sum_k \varphi(k) \int_0^1 f(x) e^{-2\pi i k x}\,dx \\ &= \sum_k c_k \varphi(k) = < \sum_k c_k \delta_k, \varphi > \end{aligned}$$

which thus concludes the proof.

$\square$

Now to a connection between Fourier transforms and Fourier series.

Proposition 1.3.1 *For $T \in \mathcal{S}'$ with period 1*

$$\begin{aligned} \mathcal{F}T &= \sum_k c_k \delta_k \\ T &= \sum_k c_k e^{2\pi i k(\cdot)} \end{aligned}$$

holds with some numbers c_k which, when T is for example an integrable function, are the usual Fourier coefficients.

Proof: Verify the formulas! Assume now the formulas to hold, and that $T = f$, where f is integrable. Take $\varphi \in \mathcal{S}$ with $\varphi(k) = 1$ and $\varphi(x) = 0,\ |x-k| > 1/2$, and we get by Poisson's Summation Formula

$$\begin{aligned} c_k &= \mathcal{F}T(\varphi) = T(\mathcal{F}\varphi) \\ &= \int_{-\infty}^{\infty} f(x)\mathcal{F}\varphi(x)\,dx \end{aligned}$$

$$
\begin{aligned}
&= \sum_{n=-\infty}^{\infty} \int_0^1 f(x)\mathcal{F}\varphi(x+n)\,dx \\
&= \int_0^1 f(x) \sum_{n=-\infty}^{\infty} \mathcal{F}\varphi(x+n)\,dx \\
&= \int_0^1 f(x) \sum_{n=-\infty}^{\infty} \varphi(n)e^{-2\pi i n x}\,dx \\
&= \int_0^1 f(x)e^{-2\pi i k x}\,dx
\end{aligned}
$$

□

Proposition 1.3.2 (Discrete Fourier Transform: DFT) *Suppose that $\hat{T} \in \mathcal{S}'$ has period 1, and that T also is periodic. Then T has an integer period N and*

$$
\begin{aligned}
T &= \sum_{k=-\infty}^{\infty} t_k \delta_k \\
\hat{T} &= \sum_{k=-\infty}^{\infty} c_k \delta_{k/N}
\end{aligned}
$$

Both sequences have period N with the relations

$$
\begin{aligned}
Nc_k &= \sum_{l=1}^{N} t_l e^{-2\pi i k l/N} \qquad && (k = 1, ..., N) \\
t_l &= \sum_{k=1}^{N} c_k e^{2\pi i k l/N} \qquad && (l = 1, ..., N)
\end{aligned}
$$

Proof: That $\hat{T}$ has period 1 gives by the previous Proposition

$$T = \sum_{k=-\infty}^{\infty} t_k \delta_k$$

Here it is evident that if T also has a period, this must be an integer N. This implies that $t_{k+N} = t_k$ for all k and yields in its turn

$$\hat{T} = \sum_{k=-\infty}^{\infty} c_k \delta_{k/N}$$

Since $\hat{T}$ has period 1 it follows that $c_{k+N} = c_k$ for all k.

We now calculate $\hat{T}$ with the aid of the expression for T.

$$\begin{aligned}
\mathcal{F}T &= \mathcal{F}\sum_{k=-\infty}^{\infty} t_k \delta_k \\
&= \mathcal{F}\left(\sum_{k=1}^{N} t_k \sum_{l=-\infty}^{\infty} \delta_{k+lN}\right) \\
&= \sum_{k=1}^{N} t_k e^{-2\pi i k(\cdot)} N^{-1} \sum_{l=-\infty}^{\infty} \delta_{l/N} \\
&= \sum_{l=-\infty}^{\infty} \left(N^{-1}\sum_{k=1}^{N} t_k e^{-2\pi i k l/N}\right) \delta_{l/N}
\end{aligned}$$

The relation $Nc_k = \sum_{l=1}^{N} t_l e^{-2\pi i k l/N} \quad (k = 1, ..., N)$ now follows. Verify the remaining relation, and that

$$\mathcal{F}\sum_{l=-\infty}^{\infty} \delta_{lN} = N^{-1}\sum_{l=-\infty}^{\infty} \delta_{l/N}$$

□

We now return to the Poisson Summation Formula, and give an alternative proof which works unaltered in higher dimensions.

Proof of Theorem 1.4; alternative: Note that the function $\sum_k \hat{\varphi}(s+k)$ is infinitely differentiable, and has period 1. We get by Theorem 1.5 (which can be proved directly without invocation of Poisson's Summation Formula: no circular argument)

$$\begin{aligned}
\sum_k \hat{\varphi}(s+k) &= \sum_l e^{2\pi i l s} \int_0^1 e^{-2\pi i l \sigma} \sum_k \hat{\varphi}(\sigma+k)\, d\sigma \\
&= \sum_l e^{2\pi i l s} \sum_k \int_0^1 e^{-2\pi i l \sigma} \hat{\varphi}(\sigma+k)\, d\sigma \\
&= \sum_l e^{2\pi i l s} \int_{-\infty}^{\infty} e^{-2\pi i l \sigma} \hat{\varphi}(\sigma)\, d\sigma \\
&= \sum_l e^{2\pi i l s} \check{\varphi}(l) \;=\; \sum_l e^{-2\pi i l s} \varphi(l)
\end{aligned}$$

The proof is complete. □

The Sampling Theorem is next on our programme. The theorem shows the possibility to reconstruct a function defined on the whole real axis in its entirety, under certain conditions on its spectrum (Fourier transform), from knowledge of its denumerable *sample values* only.

We write $\sin \pi x/(\pi x) =: \operatorname{sinc} x$ and $1_{(-1/2,1/2)}$, where the latter denotes the *cut-off function* which takes the value 1 on $(-1/2, 1/2)$, and 0 elsewhere. Note that $\mathcal{F}1_{(-1/2,1/2)} = \operatorname{sinc}$.

Theorem 1.6 (The Sampling Theorem) *Suppose that f is a smooth function with moderate growth at infinity as in Definition 1.2.2, and with $\hat{f}(s) = 0,\ |s| \geq 1/2$.* [10] *Then*

$$\begin{aligned} f &= \operatorname{sinc} * \sum_{k=-\infty}^{\infty} f(k)\delta_k \\ &= \sum_{k=-\infty}^{\infty} f(k) \operatorname{sinc}(\cdot - k) \end{aligned}$$

holds in the subspace $\mathcal{S}_{lp} = \{\varphi \in \mathcal{S}; \hat{\varphi}(s) = 0$ for $|s| \geq 1/2\} \subset \mathcal{S}$.

Proof: The operations have been defined in Proposition 1.2.1 and 1.2.3, except for the convolution with sinc. That this convolution is legitimate will be verifiable when it is done below. We have

$$\sum_k f(k)\delta_k = f \sum_k \delta_k$$

(verify!). A Fourier transformation gives, using Poisson's Summation Formula,

$$\mathcal{F}(\sum_k f(k)\delta_k) = \mathcal{F}f * \mathcal{F}\sum_k \delta_k = \mathcal{F}f * \sum_k \delta_k = \sum_k \mathcal{F}f(\cdot - k)$$

where the last expression has period 1 and the sum reduces to exactly one term on the interval $(-1/2, 1/2)$, where it coincides with $\hat{f}(s)$. Multiplication by the cut-off function (verify legitimacy!) and an inverse Fourier transformation yields the formula.

□

REMARK 5 We have chosen the sampling interval 1. If instead the sampling interval is T and $\hat{f}(s) = 0,\ |s| \geq 1/(2T)$ then

$$f = \sum_{k=-\infty}^{\infty} f(kT) \operatorname{sinc}(\cdot/T - k)$$

which is seen by putting $g(x) = f(xT)$ and using the Sampling Theorem on the function g.

[10] f is then infinitely differentiable according to Paley & Wiener, Theorem 2.1 below.

Remark 6 In technical applications it is not possible to realize δ or sinc, and neither $\hat{f}(s) = 0,\ |s| \geq 1/2$. However, approximations are possible, more or less successful.

You could, for example, approximate δ with a function $d \in \mathcal{S}$, which is 0 outside $(-1/2, 1/2)$, non-negative, with integral 1, and with $\mathcal{F}d \neq 0$ on $(-1/2, 1/2)$. Verify that there is a function d_1 such that (on $\mathcal{S}_{lp}$, see the Sampling Theorem)

$$f = d_1 * \sum_n f(n) d(\cdot - n)$$

Remark 7 Note the complication in Bracewell's book (p. 223) with the symbol $\mathrm{III}(x) = \sum_n \delta(x-n)$: $\mathrm{III}(x/\tau)$ is interpreted there ($\tau > 0$) as $\tau \sum_n \delta(x - n\tau)$ which then should be the same as $\sum_n \delta(x/\tau - n)$ (the origin of this lies in the following formula in Bracewell's book "$\delta(x/\tau) = \tau\delta(x)$" which in turn derives from the misleading 'formula' "$\int \delta(x) f(x)\, dx = f(0)$").

Verify as a contrast the following scaling of Poisson's Summation Formula

$$\sum_n \mathcal{F}\varphi(n\tau) = |\tau|^{-1} \sum_n \varphi(n/\tau)$$

and write it with δ's.

Generally, changes of variable for distributions are somewhat intricate, which might be guessed from the fact that the distributions are not defined pointwise like functions. However, translations and scalings present no special problems, as we have seen.

Example 15 (Alias effect) Assume that $\varphi \in \mathcal{S}$ with

$$\hat{\varphi}(s) = 0 \quad (|s| \geq 1) \quad \& \quad |\hat{\varphi}(s)| < \epsilon \quad (1/2 < |s| < 1)$$

This small sideband at $1/2 < |s| < 1$ outside the main allowed passband creates an error, the *alias effect*, in the following approximations.

$$\begin{aligned} \hat{\varphi}(s) &\approx \sum_{n=-\infty}^{\infty} \varphi(n) e^{-2\pi i n s} \qquad (|s| < 1/2) \\ \varphi(x) &\approx \int_{-1/2}^{1/2} \left[\sum_{n=-\infty}^{\infty} \varphi(n) e^{-2\pi i n s} \right] e^{2\pi i s x}\, ds \end{aligned}$$

which may be estimated in terms of ϵ. Estimate the maximal error as an exercise!

Example 16 A direct example of the alias effect is provided by the function $f(x) = \sin \pi x$ for which the sample values at the integers $f(n) = 0$ for all n. (f has its frequency content exactly at the critical limit $1/2$: $\mathcal{F}f = (\delta_{1/2} - \delta_{-1/2})/(2i)$.) The functions $f_k(x) = \sin k\pi x$ ($k \neq 1$ integer) all have the value 0 at the integer points. The function $f = f_1$ has then infinitely many aliases for when sampling at the integers: f_k med $k \neq 1$ ett heltal.

Differently put: given any function $f(x) = \sin a\pi x$ with $a \geq 1/2$, there is a number b with $|b| < 1/2$ such that the function $g(x) = \sin b\pi x$ coincides with f at all integer points. (Verify!)

2 Analytic Continuation

Here we will discuss things which require some theory for functions of a complex variable: the Paley-Wiener Theorem, the relation between the Laplace and the Fourier transforms, and the problem of spectral factorization.

2.1 Paley & Wiener

The feature of Paley-Wiener's Theorem is the absence of high frequencies in the spectrum, which is characterized by regularity and specific conditions on the growth of the function.

The theorem implies, among other things, that if a signal is ideally band-pass filtered then it cannot be entirely localized to finite time interval (and conversely).[11]

Theorem 2.1 (Paley & Wiener) *Let $f \in \mathcal{S}$. Then ($A > 0$)*

$$\hat{f}(s) = 0,\ |s| \geq A$$
$$\Longleftrightarrow$$
$$\begin{cases} f(x+iy) & \textit{entire} \\ |f(x+iy)| \leq C_N (1+x^2+y^2)^{-N} e^{2\pi A|y|} & \textit{for all } N \in \mathbf{N} \end{cases}$$

Remark 1 The theorem is valid also for $f \in \mathcal{S}'$ and *some* integer N; the proof then becomes more technically involved.

Proof: Suppose that $\hat{f} = 0,\ |s| \geq A$. We have

$$\begin{aligned} f(x) &= \int_{-A}^{A} e^{2\pi i x s} \hat{f}(s)\, ds \\ &= \int_{-A}^{A} \sum_{n\geq 0} \frac{(2\pi i x s)^n}{n!} \hat{f}(s)\, ds \\ &= \sum_{n\geq 0} \frac{(2\pi i x)^n}{n!} \int_{-A}^{A} s^n \hat{f}(s)\, ds \end{aligned}$$

(uniform convergence). Since $|\int_{-A}^{A} s^n \hat{f}(s)\, ds| \leq C A^{n+1}$ the radius of convergence is infinite, and f can be continued to an analytic function in the entire complex plane (f is an *entire function*).

Further we get ($(x+iy)^{2N} f(x+iy)$ also entire)

$$\begin{aligned} |(x+iy)^{2N} f(x+iy)| &= (2\pi)^{-2N} \left| \int_{-A}^{A} e^{2\pi i (x+iy)s} D^{2N} \hat{f}(s)\, ds \right| \\ &\leq C_N e^{2\pi A|y|} \end{aligned}$$

[11] The theorem is further used (through the *Support Theorem*) in the theory of partial differential equations.

which implies the desired inequality (verify!).

Conversely, assume that $f(x+iy)$ is entire with

$$|f(x+iy)| \le C(1+x^2+y^2)^{-1}e^{2\pi A|y|}$$

Let $s > A$ and Cauchy's Integral Theorem gives for $y < 0$

$$\begin{aligned}\hat{f}(s) &= \int_{-\infty}^{\infty} e^{-2\pi ixs} f(x)\, dx \\ &= \int_{-\infty}^{\infty} e^{-2\pi i(x+iy)s} f(x+iy)\, dx \\ &= \int_{-\infty}^{\infty} e^{-2\pi ixs} e^{2\pi ys} f(x+iy)\, dx\end{aligned}$$

(verify!). But $(y < 0)$

$$e^{2\pi ys}|f(x+iy)| \le Ce^{2\pi y(s-A)}(1+x^2)^{-1}$$

which implies $\hat{f}(s) = 0$, $s > A$. (Verify the case $s < -A$!) The proof is complete.

□

2.2 The Fourier-Laplace Transform

Now to the relation between Laplace and Fourier transforms. Consider $g \in \mathcal{S}$ with $gH = f$ (f *causal*). Then

$$\mathcal{F}f(s) = \int_0^{\infty} e^{-2\pi isx} f(x)\, dx$$

can clearly be continued to an analytic function of the *complex* variable s in the lower half-plane $\Im s < 0$.

Consider in particular $p = 2\pi is$, $\Re p > 0$:

$$\mathcal{F}f(\frac{p}{2\pi i}) = \int_0^{\infty} e^{-px} f(x)\, dx$$

Obviously, the right hand side is the one-sided Laplace transform of f. It is now clear that the Laplace and the Fourier transforms determine each other completely through the change of variable $2\pi is = p$. *Which transform to use is thus in principle an immaterial question.* However, established practice in engineering disciplines often makes a distinct choice depending on the application. In mathematical literature there is also a common name, the *Fourier-Laplace Transform.*

The one-sided Laplace transform of $f \in \mathcal{S}$ may be written as the two-sided Laplace transform of Hf :

$$\int_{-\infty}^{\infty} e^{-px} H(x) f(x)\, dx = \int_0^{\infty} e^{-px} f(x)\, dx = \mathcal{F}(Hf)(\frac{p}{2\pi i})$$

where we might have $f \neq Hf$. Results for the Fourier transform can now be directly translated to results for the Laplace transform and *vice versa.* (Usually, the integral in the two-sided Laplace transform converges in a vertical strip in the complex plane, $a < \Re p < b$.)

EXAMPLE 1 If we denote the two-sided Laplace transform of $f \in \mathcal{S}$, when the definition has meaning, by $\mathcal{L}f$ (the one-sided will thus be $\mathcal{L}(Hf)$) then we get $(2\pi is = p)$

$$\begin{aligned}\mathcal{L}(HDf) &= \mathcal{F}(HDf) = \mathcal{F}\{D(Hf) - f(0)\delta\} = 2\pi i(\cdot)\mathcal{F}(Hf) - f(0) \\ &= (\cdot)\mathcal{L}(Hf) - f(0)\end{aligned}$$

(recall that $D(Hf) = f\delta + HDf;\ f\delta = f(0)\delta$).

This is the one-sided Laplace transform of the derivative of a function f expressed in the same transform of the function and its value at 0.

We now show with the *potential method* how initial value problems may be treated with the Fourier-Laplace transform.

EXAMPLE 2 (Bracewell, page 394) Consider the initial value problem

$$\begin{cases} y'' + 3y' + 2y = 2 \\ y(0) = 1,\ y'(0) = 0 \end{cases}$$

First, we consider a general right hand side $S \in \mathcal{S}'$, which is a function with possibly a term $\sum_{k=0}^{N} a_k \delta^{(k)}$ added. A *fundamental solution* (Green's function, potential function, impulse response)[12] may be constructed, which yields the solution for a general right-hand side S.

Consider, with such a right-hand side S, the initial value problem

$$\begin{cases} y'' + 3y' + 2y = S \\ y(0+) = 1,\ y'(0+) = 0 \end{cases}$$

The solution will be a tempered distribution which, apart from possibly a sum of δ and its derivatives, is a function (verify!).[13]

The fundamental solution G is defined as the solution to the equation

$$G'' + 3G' + 2G = \delta \quad \text{with} \quad G(x) = 0,\ x < 0$$

G may be obtained by a Fourier transform, a partial fraction decomposition, and an inverse Fourier transformation (verify!). this gives $G(x) = (e^{-x} - e^{-2x})H(x)$ here — draw the graph and realize why!

[12] These are treated in courses on differential equations.

[13] $y(0+)$ *etc.* denote limits of the function part. (Verify that these exist!)

With $z = G * S$ we get[14]

$$z'' + 3z' + 2z = S \quad \text{and} \quad z(0+) = a,\ z'(0+) = b$$

where a, b depend on S. (Apart from possibly δ and its derivatives, z becomes a function.) When $S = 2$, we have $z = 1$.

Adding z and a suitable solution to the homogeneous equation, the desired solution is obtained

$$y = z + c_1 e^{-(\cdot)} + c_2 e^{-2(\cdot)}$$

The constants c_1, c_2 are chosen such that $y(0+) = 1,\ y'(0+) = 0$, which is possible (verify) since the solutions $e^{-(\cdot)},\ e^{-2(\cdot)}$, to the homogeneous equation are linearly independent. When $S = 2$, we get $c_1 = c_2 = 0$.

In the case $S = 2$, we may alternatively put $T = Hy$; y is continuously differentiable here. The equation becomes (compare Example 1)

$$T'' + 3T' + 2T = 2H + 3\delta + \delta'$$

where the initial conditions now are incorporated into the right-hand side. The solution is obtained by a Fourier transformation, simplification, and an inverse Fourier transformation. It is $T = H$, which gives $y = 1$.

2.3 Spectral Factorization

We conclude with the problem of spectral factorization: assume that we have observed the *energy spectrum,* $|\mathcal{F}f|^2$, of, say, an electrical signal $f \in \mathcal{S}$. You might think that f has the dimension Volt and the variable the dimension second. Parseval's Formula

$$\int |f(x)|^2\, dx = \int |\mathcal{F}f(s)|^2\, ds$$

expresses then the total energy of the signal in two ways: $|f(x)|^2$ is the energy density in time, while $|\mathcal{F}f(s)|^2$ is the energy density in frequency, *energy spectrum.*[15]

The problem of spectral factorization is, given the energy spectrum$|\mathcal{F}f|^2$, to find the function f. Consequently, all information about the phase of $\mathcal{F}f$ is missing.

The problem is, of course, not uniquely solvable. Clearly,

$$|\mathcal{F}f(s)|^2 = |e^{i\theta(s)}\mathcal{F}f(s)|^2$$

that is, multiplicating the Fourier transform with a phase factor $e^{i\theta(s)}$ does not change the energy spectrum.

[14] Note that $z = G * S$ is a particular solution as defined in elementary analysis courses.

[15] The term *power spectrum* is used somewhat differently in connection with stationary stochastic processes.

In applications, it is not unusual that the function sought is causal. If we know that f is causal ($f = Hf$) then the solution to the spectral factorization problem is unique up to a constant factor of modulus 1.[16] This is a consequence of the representation for causal f

$$\mathcal{F}f(s) = \int_0^\infty e^{-2\pi isx} f(x)\, dx$$

which can be continued to an analytic function in the lower half-plane $\Im s < 0$. Two analytic functions with the same modulus differ with at most a constant factor of modulus 1.

[16] If f is real-valued the solution is unique up to a change of sign.

3 Two Probability Theorems

We will give a simple proof of the Central Limit Theorem, and then describe the connection between autocorrelation functions and probability measures.

3.1 The Central Limit Theorem

The probability measure corresponding to the sum of n independent stochastic variables with two equally likely outcomes may, normalized to mean value 0 and variance 1, be represented by the convolution (the number of factors is n)

$$T_n = (\frac{1}{2}\,\delta_{-1/\sqrt{n}} + \frac{1}{2}\,\delta_{1/\sqrt{n}}) * \cdots * (\frac{1}{2}\,\delta_{-1/\sqrt{n}} + \frac{1}{2}\,\delta_{1/\sqrt{n}})$$

Theorem 3.1 (The Central Limit Theorem) *With T_n as above, we have (in $\mathcal{S}'$)*

$$\lim_{n\to\infty} T_n = \frac{1}{\sqrt{2\pi}} e^{-(\cdot)^2/2}$$

Proof: We have (in $\mathcal{S}'$)

$$\lim_{n\to\infty} \mathcal{F}T_n = \lim_{n\to\infty} (\cos\frac{2\pi(\cdot)}{\sqrt{n}})^n = e^{-2\pi^2(\cdot)^2}$$

(verify the last elementary limit!) An inverse Fourier transformation gives the result.

□

Remark 1 The same proof may be used for the case with an arbitrary probability measure with finite variance (one such is the above $\frac{1}{2}\,\delta_{-1} + \frac{1}{2}\,\delta_1$).

3.2 Autocorrelation Functions

Definition 3.2.1 *The* ***autocorrelation function*** *of a function $\varphi \in \mathcal{S}$ is written $\varphi \star \varphi$ and is defined by*

$$\varphi \star \varphi(x) := \int \varphi(x+u)\varphi^*(u)\,du$$

Note that

$$\varphi \star \varphi = \varphi * (\varphi^*)\check{}$$

($\check{\varphi}(x) := \varphi(-x)$). Furthermore, we have ($\varphi \in \mathcal{S}$)

$$\mathcal{F}\{\varphi \star \varphi\} = |\mathcal{F}\varphi|^2$$

(verify!) Every such autocorrelation function has thus a non-negative Fourier transform.[17]

We will now (Theorem 3.2 and Remark 3) give an answer to the question: *Which objects have (like the autocorrelation function $\varphi \star \varphi$) Fourier transforms that are bounded positive measures?* It is precisely these measures which can be normalized to probability measures, if divided by the *total mass*: the least constant C in the following definition.

Definition 3.2.2 *Let $T \in \mathcal{S}'$ and $\varphi \in \mathcal{S}$.*

*T is called a **tempered measure** if, for all $\varphi \in \mathcal{S}$ with $\varphi = 0$ outside a fixed bounded interval,*

$$|T(\varphi)| \leq C \sup_x |\varphi(x)|$$

*holds. If C is independent of the interval, the measure is called **bounded**.*

*T is called **positive** if $T(\varphi) \geq 0$ for all $\varphi \geq 0$, and we write $T \geq 0$.*

*T is called **positive definite** if $T(\varphi \star \varphi) \geq 0$ for all φ.*

A motivation for the term 'positive definite' appears in Example 2 below.

Proposition 3.2.1 *Let $T \in \mathcal{S}'$. Then $\mathcal{F}T$ is positive if, and only if, T is positive definite.*

In addition, if T is positive then T is a (positive tempered) measure.

Proof: The following hold for all $\varphi \in \mathcal{S}$

$$\mathcal{F}T(\varphi) \geq 0,\ \varphi \geq 0 \Leftrightarrow \mathcal{F}T(|\varphi|^2) \geq 0 \Leftrightarrow T(\varphi \star \varphi) \geq 0$$

(verify the last step!) The first step (in the non-trivial direction) is verified by putting, for $0 \leq \psi \in \mathcal{S}$ with the value 0 outside a bounded interval,

$$\psi_n(x) = (\psi(x) + e^{-x^2}/n)^{1/2}$$

and noting that ψ_n^2 tends to ψ in $\mathcal{S}$ (verify!).

To prove the second statement, take real-valued *(sic!)* $\varphi_n \in \mathcal{S}$ which are 0 outside a fixed bounded interval and with $\sup_x |\varphi_n(x)| \to 0$. Let $0 \leq \varphi \in \mathcal{S}$ be 1 on the interval, and take $\varepsilon > 0$ arbitrarily. For n large enough, $\varepsilon\varphi \pm \varphi_n \geq 0$ holds, which yields $\varepsilon T(\varphi) \pm T(\varphi_n) \geq 0$, or $|T(\varphi_n)| \leq \varepsilon T(\varphi)$. We get $T(\varphi_n) \to 0$, which implies the desired inequality (verify!). The proof is complete. □

[17] $\varphi \star \varphi$ is sometimes normalized through a division by the scaling factor $\int |\varphi(u)|^2\, du$, and is then written γ. Then $\int \hat{\gamma}(s)\, ds = 1$ and $\check{\gamma} \geq 0$; whence *probability measure.*

EXAMPLE 1 If $\varphi \in \mathcal{S}$ then $\varphi \star \varphi$ is positive definite (verify!).

EXAMPLE 2 Suppose f is a continuous function, and also a positive definite tempered distribution. Then we have

$$|f(x)| \leq f(0) \quad \text{and} \quad \check{f} = f^*$$

We verify this by writing

$$0 \leq \int f(x)\varphi \star \varphi(x)\, dx = \int\int f(x+y)\varphi(x)\varphi(-y)^*\, dxdy$$

and choosing φ there which approximate $\sum_{j=1}^{n} z_j \delta_{x_j}$. This gives the condition

$$\sum_{j,k} f(x_j - x_k) z_j z_k^* \geq 0$$

which is satisfied for all choices of $x_j,\ z_j$.

In the case $x_1 = 0,\ x_2 = x$, the condition implies that the matrix

$$\begin{bmatrix} f(0) & f(x) \\ f(-x) & f(0) \end{bmatrix}$$

is Hermitian and positive (verify!). In particular, we get $\check{f} = f^*$ and $|f(x)| \leq f(0)$, which we wanted to prove.

Now the the matrix $[f(x_j - x_k)]$ is clearly Hermitian and positive. That this matrix is *positive definite* means by definition (of positive definite matrices) that equality in the condition is attained only when all $z_j = 0$.[18]

EXAMPLE 3 Let $T \in \mathcal{S}'$ be positive definite. Then $\check{T} = T^*$ holds, where $T^*(\varphi) := T(\varphi^*)^*$. Verify this!

Verify also that δ is a positive bounded measure with total mass 1 (as well as any non-negative integrable function with integral 1). Moreover, verify that the function $f(x) = x^2$ is a positive unbounded measure.

Check finally that, *e.g.*, the function $f(x) = e^{x^2}$, which does not belong to $\mathcal{S}'$, still enjoys the first property in Definition 2. The function f is a positive measure which is not tempered. We will not go into further details here.

Theorem 3.2 (Bochner) *Suppose $f \in \mathcal{S}'$ is positive definite and a continuous function. Then $\mathcal{F}f$ is a bounded positive measure.*

[18] It can be shown that (verify!), for a positive definite continuous function $f \in \mathcal{S}'$, if $|f(x_0)| = f(0)$ and $|f(x)| < f(0),\ 0 < x < x_0$, then $\mathcal{F}f = \sum_k a_k \delta_{(k+\alpha)/x_0}$ where $\sum_k a_k = f(0),\ a_k \geq 0$, and $f(x_0) = e^{2\pi i \alpha} f(0)$. In this case the matrix need not be positive definite (verify!).

Proof: According to Proposition 3.2.1, $\mathcal{F}f \geq 0$ and is a positive measure. The continuity of f will now yield finite total mass of $\mathcal{F}f$. Choose $0 \leq \varphi \in \mathcal{S}$ such that $\varphi(0) = 1$ and $\mathcal{F}\varphi(s) = 0, \ |s| \geq 1$ (verify that this is possible!). Put $\varphi_n(x) = \varphi(x/n), \ n = 1, 2, \dots$.

We get ($\mathcal{F}f \geq 0$, f continuous)

$$0 \leq \mathcal{F}f(\varphi_n) = f(\mathcal{F}\varphi_n) = \int_{|x|\leq 1/n} f(x)\mathcal{F}\varphi_n(x)\, dx \longrightarrow f(0)$$

Consider now an arbitrary bounded interval (a, b), and observe that $\varphi_n(x) \geq 1 - \varepsilon$ holds in the interval if n is sufficiently large. Take $\psi \in \mathcal{S}$ that is 0 outside the interval (a, b). Then ($\varepsilon > 0$ arbitrary)

$$\varphi_n(x) \pm (1-\varepsilon)\psi(x)/\sup_x |\psi(x)| \geq 0$$

follows, and thus[19]

$$(1-\varepsilon)|\mathcal{F}f(\psi)|/\sup_x |\psi(x)| \leq \mathcal{F}f(\varphi_n) \leq f(0) + \varepsilon$$

if n is large enough. Thus we have

$$|\mathcal{F}f(\psi)| \leq f(0)\sup_x |\psi(x)|$$

The proof is complete. □

REMARK 2 Using the more general concept of distribution (not only tempered ones) and the corresponding definition of positive definite distribution, Schwartz' Theorem holds: *T is positive definite precisely when $\mathcal{F}T$ is a positive measure.* An idea for a proof is to regularize T by convolving it with approximate δ to a continuous function, then use Bochner's Theorem 3.2, and take limits.

In the tempered case, this is Proposition 3.2.1. There the positive measure is a tempered distribution, which narrows the possibilities (see Example 3).

REMARK 3 For a positive measure with finite total mass, it can be shown that its Fourier transform is a (positive definite) continuous function, the value of which at 0 is the total mass of the measure.

[19] Compare to the proof of Proposition 3.2.1.

4 Selected Landings

We will land at selected places in Bracewell's book, in about the order things appear there. We start with the Uncertainty Relation, treat then Gibb's Phenomenon, followed by the Radon Transform. Our next landing is in antennas and thin lenses. We conclude with a discussion of some issues concerning discretization and Fourier transform.

4.1 The Uncertainty Relation

The Uncertainty Relation in quantum mechanics is mathematically the fact that the two integrals

$$\int |Df(x)|^2 \, dx \; (= 4\pi^2 \int |s\mathcal{F}f(s)|^2 \, ds) \quad \text{and} \quad \int |xf(x)|^2 \, dx$$

cannot both be small jointly. This is quantitatively expressed by the following theorem.

Theorem 4.1 (The Uncertainty Relation) *Let $f \in \mathcal{S}$ with $\int |f(x)|^2 \, dx = 1$. Then*

$$\frac{1}{2} \leq (\int |Df(x)|^2 \, dx)^{1/2} (\int |xf(x)|^2 \, dx)^{1/2}$$

holds.

Proof: The following identity is the foundation of the proof. (Verify the identity!)

$$f = D\{(\cdot)f\} - (\cdot)Df$$

We now use the identity, a partial integration, and the Cauchy-Schwarz' inequality.

$$\begin{aligned} 1 &= \int f(x)f(x)^* \, dx = \int D\{xf(x)\}f(x)^* \, dx - \int xDf(x)f(x)^* \, dx \\ &= -\int xf(x)Df(x)^* \, dx - \int Df(x)xf(x)^* \, dx \\ &\leq 2(\int |Df(x)|^2 \, dx)^{1/2} (\int |xf(x)|^2 \, dx)^{1/2} \end{aligned}$$

The proof is complete.

$\square$

4.2 Gibbs' Phenomenon

The partial sums of a Fourier series belonging to a function with a jump discontinuity all display an overshoot close to that point. This is called *Gibbs' Phenomenon.*

Let f and g have period 1, and let $D^2 g$ be continuous. Then the Fourier series of g converges uniformly to g at all points (Theorem 1.5). Assume that

$$g\, 1_{(-1/2,1/2)} = (f - (H - 1/2))\, 1_{(-1/2,1/2)}$$

This means that f has a *unit jump at the integers and at the half-integers* compared to g; otherwise they have the same regularity. (Draw a picture!)

We will now investigate $f - g$ which is a square wave — Gibb's Phenomenon for f will be the same as for the square wave $f - g$ (verify!). In the interval $(0, 1/2)$ we consider the difference (which produces the overshoot)

$$\sum_{n=-N}^{N} c_n e^{2\pi i n x} - (H(x) - 1/2) = -\sum_{|n|>N} c_n e^{2\pi i n x}$$

(this may be shown to converge pointwise in $0 < |x| < 1/2$).

Calculating the coefficients c_n, we have (with $N = 2M + 1$)

$$\sum_{n=-N}^{N} c_n e^{2\pi i n x} = \sum_{k=0}^{M} \frac{2}{\pi(2k+1)} \sin 2\pi(2k+1)x$$

The smallest positive extremum point (put the derivative equal to 0) is here $x = 1/(4M + 4) = 1/(2N + 2)$, which gives

$$-\sum_{|n|>N} c_n e^{\pi i n/(N+1)} = -\sum_{k=M+1}^{\infty} \frac{2}{\pi(2k+1)} \sin(\pi(2k+1)/(2M+2))$$

When $N \to \infty$, it follows that $M \to \infty$ and the right-hand side (which is a Riemann sum) converges to

$$-\int_1^{\infty} \frac{\sin \pi x}{\pi x}\, dx \;\approx\; 0,0894899$$

(where the value has been computed by MATHEMATICA).

The overshoot is thus about 9% of the jump as $N \to \infty$.

4.3 The Radon Transform

In dimension 2, the Radon transform of a function is its integral over all lines. This transform is used, for example, in Computer Tomography (CT), in Magnetic Resonance Imaging (MRI), in Positron Emission Scanning (PET), and in Synthetic Aperture Radar (SAR).[20]

[20] In SAR, the lines are replaced by circles.

In Computer Tomography, the function $f(x)$ represents the absorption coefficient in the material (tissue) per unit length, and the absorption is observed for X-rays traversing a cross-section of the object (body) along lines L

$$\int_L f(x(l))\, dl$$

This is now in theory recorded for all lines L, and the task is to reproduce the function values $f(x)$ from the values of all the line integrals.[21]

Definition 4.3.1 *The **Radon Transform** of a function $f \in \mathcal{S}$ is defined by*

$$\mathcal{R}_\theta f(s) := \int_{x\cdot\theta=s} f(x)\, dx$$

where $|\theta| = 1$.[22]

Note that $\mathcal{R}_\theta f(s) = \mathcal{R}_{-\theta} f(-s)$, and that the requirement $|\theta| = 1$ is made to have the line correspond bijectively, apart from a sign, to (θ, s).

REMARK 1 The *Abel Transform* of the *radial* function f in dimension 2 is defined by $((x > 0))$

$$\mathcal{A}f(x) := 2\int_{r>x} f(r)\frac{r\, dr}{\sqrt{r^2 - x^2}}$$

and $\mathcal{A}f(-x) := \mathcal{A}f(x)$.[23]

Verify that the Radon transform of a radial function coincides with its Abel transform. This means that, *for radial functions in dimension 2, composing an Abel transform with a Hankel transform*[24] *is the same as composing the Radon transform and the Fourier transform.*

REMARK 2 In dimension $n \geq 3$, the integration may be done in more than one way. One is to integrate over the $(n-1)$-dimensional (hyper-)plane $x \cdot \theta = s$; another is to integrate over the lines $x = t + l\eta$, where $|\eta| = 1$ and $t \cdot \eta = 0$. In applications, integration over lines is commonly used – images in dimension 3 are often built from plane slices, the latter being reconstructed from line integrals.

A natural question might now be: Which families of 'surfaces' or 'lines' are admissible for a reconstruction of a function from values of its integrals over these to be possible?[25]

---

[21] The viability of this task was shown by Radon about a century ago.

[22] The variable s has a different role in this section, and θ denotes a point on the unit circle (sphere)!

[23] The last statement is used, but not explicitly made in Bracewell's book.

[24] The Fourier transform of a radial function is the Hankel Transform.

[25] We refer to the book HELGASON S., *The Radon Transform*, Birkhäuser, 1980, where a comprehensive treatment of the central issues may be found.

REMARK 3 In implementations, problems arising from discretization, sampling, and reconstruction, will arise.[26] One such is that the Radon transform is expressed in polar coordinates, while the reconstruction is done with the Fast Fourier Transform in rectangular coordinates ...

We now show the theoretical result on which all of the techniques CT, PET, MRI, and SAR are based.

Theorem 4.2 (Radon) *Let $f \in \mathcal{S}$. Then*

$$\mathcal{F}\mathcal{R}_\theta f(\sigma) = \mathcal{F}f(\sigma\theta)$$

holds.

Proof: It suffices to consider the case $\theta = (1,0)$, since a rotation of the coordinate system in the (x_1, x_2)-plane corresponds to the same rotation in the Fourier domain. (Verify!)

We get, with $\theta = (1,0)$, that $x \cdot \theta = x_1$, and thus

$$\begin{aligned}
\mathcal{F}\mathcal{R}_\theta f(\sigma) &= \int_{-\infty}^{\infty} e^{-2\pi i\sigma s} \int_{x_1 = s} f(x_1, x_2)\, dx_2 ds \\
&= \int\int e^{-2\pi i\sigma(1,0)\cdot(x_1,x_2)} f(x_1, x_2)\, dx_1 dx_2 \\
&= \mathcal{F}f(\sigma\theta)
\end{aligned}$$

□

We end with a starting point for reconstruction of a function from its Radon transform. Recall that the Hilbert transform corresponds to multiplication by $i\,\mathrm{sign}\,(\cdot)$ in the Fourier domain, and that differentiation corresponds to multiplication by $2\pi i(\cdot)$.

EXAMPLE 1 Let $f \in \mathcal{S}$. Then, in dimension 2 after a change to polar coordinates, $(\mathcal{R}_\theta f(s) = \mathcal{R}_{-\theta} f(-s))$

$$\begin{aligned}
f(x) &= \int e^{2\pi i x\cdot\xi} \mathcal{F}f(\xi)\, d\xi \\
&= \int e^{2\pi i\sigma\theta\cdot x} \mathcal{F}f(\sigma\theta)\sigma\, d\sigma d\theta \\
&= 1/2 \int e^{2\pi i\sigma\theta\cdot x} \mathcal{F}\mathcal{R}_\theta f(\sigma)\sigma\,\mathrm{sign}\,\sigma\, d\sigma d\theta \\
&= -1/(4\pi) \int e^{2\pi i\sigma(\theta\cdot x - s)} i\,\mathrm{sign}\,\sigma\; 2\pi i\sigma\mathcal{R}_\theta f(s)\, ds d\sigma d\theta
\end{aligned}$$

where the integration, from the third equality sign on, is made over all real σ, *i.e.*, also over negative values.

[26] Further information on these matters is available in the book NATTERER F., *The Mathematics of Computerized Tomography*, Wiley, 1986.

4.4 Antennas and Thin Lenses

In this section, we discuss *coherent* electromagnetic radiation: the wavelength (and the frequency) is thus fixed throughout.

First, we consider the relation between the aperture field and the direction characteristics of an antenna, which is approximately given by the Fourier transform.

Second, we will argue that a field in one focal plane of a thin convex lens creates approximately its Fourier transform in the opposite focal plane.

Antennas

Here we just briefly reiterate the argument in Bracewell's book, and use the same notation.

Consider the case when the field in the aperture of the antenna may be described by one position variable only: $E(x)e^{i\omega t}$, where ω is the circular frequency. At the point P at the distance r from the point x, the contribution to the far field will be $E(x)e^{i\omega t}e^{-2\pi ir/\lambda}$ from the aperture field by Huyghens' Principle, where $\lambda\omega/(2\pi)$ is the field propagation velocity. Let now R denote the distance between the point $x = 0$ and the point P, and θ the angle between the horizontal axis and the line through $x = 0$ and P.

The Cosine Theorem gives

$$r^2 = R^2 + x^2 - 2xR\cos(\theta + \pi/2)$$

or

$$r = R\left(1 + 2(x/R)\sin\theta + (x/R)^2\right)^{1/2}$$

For $x \ll R$ (far away compared to the antenna dimensions), we approximately have

$$r = R + x\sin\theta$$

which gives, after integration over x and with $s = (\sin\theta)/\lambda$, the field at P

$$e^{-2\pi iR/\lambda + i\omega t}\int_{-\infty}^{\infty} E(x)e^{-2\pi ixs}\,dx = e^{-2\pi iR/\lambda + i\omega t}\hat{E}(s)$$

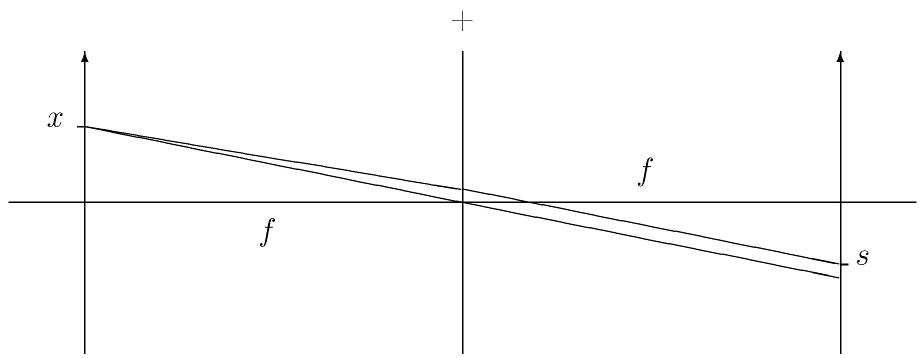
Since $s = (\sin\theta)/\lambda$, this is essentially the direction characteristics as $|\theta| \ll 1$.

Example 2 For $E = 1_{(-1/2,1/2)}$, δ, $\delta_{-1/2} + \delta_{1/2}$ The characteristics will be approximatively respectively sinc, 1, $\cos(\pi(\cdot)/\lambda)$.

A Thin Convex Lens

We will consider a thin convex lens with focal distance f and, in one focal plane, the field $E(x)e^{i\omega t}$. We will restrict our discussion to the geometrical optics approximation and 'central rays'. Notation will found in the figure below.

Suppose thus that the field in one focal plane is given by $E(x)e^{i\omega t}$, where ω is the circular frequency, and $\lambda\omega/(2\pi)$ is the field propagation velocity.



In the geometrical optics approximation, rays from a point x in one focal plane is refracted to parallel rays by the lens, and the ray through the center is not refracted. This implies that a ray from the point x to the point s has travel length

$$\left(f^2+|s|^2\right)^{1/2}+\left(f^2+|x|^2\right)^{1/2}$$

which is, when $|x| \ll f,\ |s| \ll f$, approximatively $2f+(|x|^2+|s|^2)/(2f)$. The assumption about central rays implies $|x| \approx |s|$, and so the approximate travel length amounts to

$$2f+|x||s|/f$$

By Huyghens' Principle, the contribution from the point x to the field at the point s is thus, since $xs=-|x||s|$ from the geometrical optics approximation, $E(x)e^{i\omega t}e^{-2\pi i(2f-xs/f)/\lambda}$. This yields, after an integration over x, the whole field at the point s

$$e^{-4\pi if/\lambda+i\omega t}\int E(x)e^{2\pi ixs/(f\lambda)}\,dx=e^{-4\pi if/\lambda+i\omega t}\hat{E}(-s/(f\lambda))$$

This may be expressed as *the field in one focal plane generates its Fourier transform in the other focal plane.*

Example 3 Let the function

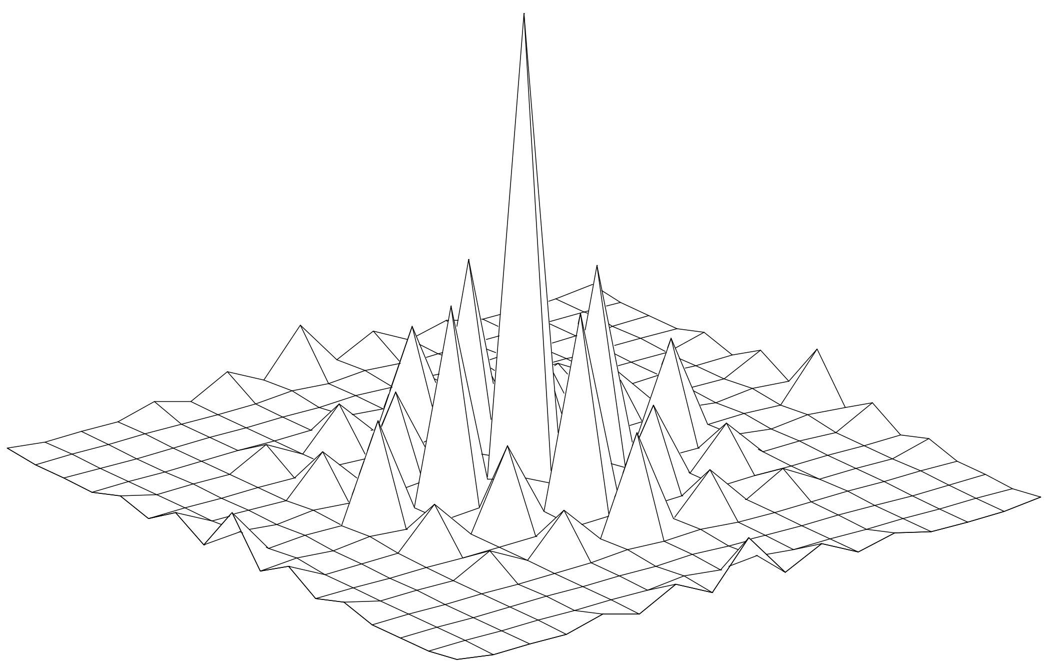
$$f = \left[\left(1_{(-1/2,1/2)}((\cdot)_1)/\epsilon)/\epsilon \, 1_{(-1/2,1/2)}((\cdot)_2/\epsilon)/\epsilon \right) * \sum_n \delta_n \right] 1_{(0,1)}(|\cdot|/R)/R^2$$

represent an infinite square lattice. The squares have edge-length ϵ and are centered at the integer points $n = (n_1, n_2)$. The lattice is circularly cut off with an iris diaphragm of radius R.

A Fourier transform gives

$$\begin{aligned} \hat{f} &= \left[\operatorname{sinc}(\epsilon(\cdot)_1) \operatorname{sinc}(\epsilon(\cdot)_2) \sum_n \delta_n \right] * J_1(2\pi R|\cdot|)/(R|\cdot|) \\ &= \sum_n \operatorname{sinc}(\epsilon n_1) \operatorname{sinc}(\epsilon n_2) J_1(2\pi R|(\cdot) - n|)/|(\cdot) - n| \end{aligned}$$

The graph of $|\hat{f}|^2$ below has been created by MatLab. The computation required about 40 Mflop to produce the graph from the 289 function values used. The parameters were $\epsilon = 2/3$, $2\pi R = 20$, $-4 < s_1 < 4$, $-4 < s_2 < 4$. (The intervals were increased 0.001 at the boundary points.)



4.5 Some Issues of Discretization

Sampling and Fourier Transformation

A numerical treatment of functions and Fourier transforms is commonly preceded by an approximation. For example, a function $f \in \mathcal{S}$ may be approximated by a finite number of values, each value being represented by a finite decimal expansion.[27]

We will assume that a low-pass filtering has been performed, so that $\mathcal{F}f(s) = 0,\ |s| \geq 1/2$. The Sampling Theorem then tells that the sample values $f(n),\ n \in \mathbf{Z}$, contain the information about all other function values, and ($|s| < 1/2$)

$$\mathcal{F}f(s) = \sum_n f(n)e^{-2\pi ins} = \sum_n f(n)z^{-n} = F(z)$$ [28]

In the example chosen here, $f \in \mathcal{S}$, in general there are infinitely many non-zero sample values. Then all but a finite number, N say, have to be discarded. We let, with A as in Approximation,

$$\mathcal{F}f_A(s) := \sum_{n=0}^{N-1} f(n)e^{-2\pi ins}$$

Put $F_\nu := N^{-1}\mathcal{F}f_A(\nu/N),\ f_n := f(n)$, and we get the Discrete Fourier Transform, DFT. Compare Proposition 1.3.2 above!

$$\begin{aligned} NF_\nu &= \sum_{n=0}^{N-1} f_n e^{-2\pi in\nu/N} \qquad (0 \leq \nu \leq N-1) \\ f_n &= \sum_{\nu=0}^{N-1} F_\nu e^{2\pi in\nu/N} \qquad (0 \leq n \leq N-1) \end{aligned}$$

What is the error in the transform due to the approximations? A rough first estimate shows ($|s| < 1/2$)

$$|\mathcal{F}f(s) - \mathcal{F}f_A(s)| \leq \sum_{n \neq 0,\ldots,N-1} |f(n)|$$

where the error is expressed in the discarded values. The subsequent restriction to F_ν entails no additional loss of information (verify!).

[27] Round-off errors appearing when taking finite decimal expansions, *quantization errors,* are usually supposed to be independent and normally distributed. We will subsequently assume that the sample values are exact, and thus disregard quantization errors.

[28] *The z-transform of the sample sequence is the Fourier transform of the continuous time signal with $z = e^{2\pi is},\ |s| < 1/2$.*

Diverse Convolutions

The convolution of two tempered distributions which are continuous periodic functions (with the same period) cannot be given meaning within our framework. Verify this by considering the Fourier transform of a proposed convolution.

However, a *periodic convolution* of two periodic functions f och g, both with period 1, may be defined by

$$f \circledast g(x) := \int_0^1 f(x-y)g(y)\,dy$$

where $f \circledast g$ gets period 1. The periodic convolution corresponds to multiplication of the Fourier coefficients (verify!).

Note that a function which is 0 outside a bounded interval can be treated as periodic without loss of information (keeping in mind that it is not periodic but 0 outside the original interval!).

For sequences, $\{f_n\}$, $\{g_n\}$, the standard definition of convolution

$$(f * g)_n = \sum_{k=-\infty}^{\infty} f_{n-k} g_k$$

corresponds to a multiplication of their z-transforms. If the sequences have finite length M (*i.e.* the value is 0 except for M consecutive ones) and N respectively, then the convolution will, in general, have length $M + N - 1$. (Verify!)

Two sequences of length N, $\{f_n\}$ and $\{g_n\}$, may be continued periodically and then a *periodic convolution* may be defined:

$$(f \circledast g)_n := \sum_{k=0}^{N-1} f_{n-k} g_k$$

which has the same period N. For the Fourier transforms above, F_ν and G_ν, (which also may be continued with period N) this convolution again corresponds to multiplication (verify!).

The periodic convolution and the usual one coincide if the sequences are extended by zeroes so as to double their length. Let $\{f_n\}_{n=0}^{N-1}$, $\{g_n\}_{n=0}^{N-1}$ both have length N, and let $\{f_n^0\}$, $\{g_n^0\}$ be the same with $N-1$ zeroes appended at one end, say. Then

$$(f^0 \circledast g^0)_n = (f * g)_n \qquad (0 \le n \le 2N-2)$$

holds. Verify this!

Fast Fourier Transform: FFT

We describe the key idea behind a fast algorithm, FFT, to compute the Fourier transform of sequences of certain lengths N. We take $N = 2^n$, but the same idea applies in the case $N = p_1 p_2 \cdots p_n$ with p_i prime numbers. The number of operations *multiply-add* becomes $N\ {}^2\!\log N$, when $N = 2^n$.[29]

Let first $N = 2$. We have

$$\begin{aligned} X_0 &= x_0 + x_1 \\ X_1 &= x_0 - x_1 \end{aligned}$$

Clearly, 2 operations were needed.

Let next $N = 4$. We have

$$\begin{array}{rcrcrcrcr} X_0 = & x_0 & + & x_1 & + & x_2 & + & x_3 \\ X_1 = & x_0 & + & ix_1 & - & x_2 & - & ix_3 \\ X_2 = & x_0 & - & x_1 & + & x_2 & - & x_3 \\ X_3 = & x_0 & - & ix_1 & - & x_2 & + & ix_3 \end{array}$$

Here, the case $N = 2$ may be used on the 2 couples (x_0, x_2) and (x_1, x_3), and then 4 operations performed; thus a total of 8 operations.

Let finally $N = 8$. Partition the sequence into 2 groups:

$$(x_0, x_2, x_4, x_6) \quad \text{and} \quad (x_1, x_3, x_5, x_7)$$

Use the case $N = 4$ on each group. Verify that an additional 8 operations conclude the calculation! (1 per coefficient.)

Hopefully, the key idea behind the FFT is now discernible. This constitutes the main loop in the design of computer implementations for general $N = 2^n$.[30]

How many operations are needed in the general case $N = p_1 p_2 \cdots p_n$?

Example 4 Multiplication of two numbers corresponds to a convolution of their digit sequences in any base. (Compare the z-transform.)

If both numbers have $N = 2^n$ digits, a straight-forward calculation of the convolution would need N^2 multiply-add operations — this is a quadratic dependence on the number of digits. If, instead, both sequences are first treated with FFT, then pointwise multiplied, and finally given an inverse FFT, then the number of operations will be $N + 3N\ {}^2\!\log N$ — which grows considerably slower than quadratic in the number of digits.

A practical problem in the usage of FFT is that not all sequences have length $N = 2^n$. This problem, and the problem of discarding sample values, will be topics in computer exercises.

---

[29] A straight-forward calculation of the transform requires in this case $N(N-1)$ multiply-adds (verify!)

[30] A discussion of the mathematical ramifications of the main idea, and some historical material, may be found in Auslander L. & Tolimieri R., Is computing with the finite Fourier transform pure or applied mathematics? *Bull. Amer. Math. Soc.* **1**:6 (1979), 847–897. (The department library has the journal.)

5 Wavelet Analysis – a Sketch

A drawback with the Fourier transform is that a local change at one frequency influences the corresponding function globally: a change in one term of the Fourier series alters the function everywhere.

One way to circumvent this drawback is to perform a *wavelet transform*.[31]

A given function $f \in L^2$ ($|f|^2$ integrable) is decomposed into a sum

$$f(x) = \sum_{k,l=-\infty}^{\infty} a_{k,l} 2^{-k/2} \psi(2^{-k}x - l)$$

where

$$a_{k,l} = \int f(x) 2^{-k/2} \psi(2^{-k}x - l)^* \, dx$$

In more condensed notation, this becomes

$$f = \sum_{k,l=-\infty}^{\infty} a_{k,l}\psi_{k,l} \quad \text{where} \quad a_{k,l} = (f, \psi_{k,l})$$

Each term is, apart from a scale factor, a dilated and translated version of a single function, the wavelet $\psi \in L^2$, which will be mainly localized in a time interval. If a narrow frequency band is desired, the price is worsened localizaton in time, and conversely, due to the Uncertainty Relation.[32]

Note that the scale doubles when k is increased to $k + 1$ — this is the reason for the minus sign in 2^{-k}.[33]

The wavelet system may be chosen orthogonal or not, depending on the specific application.

The wavelet transform has, as the Fourier transform, both a continuous, a discrete, and a finite version. *We only discuss the discrete and finite version here.*[34]

Computationally, the wavelet algorithm is better than the FFT: the number of operations multiply-add is at most $(n + 1)\, 2^{K+1}$ when the sequence length is 2^K, where n is a constant depending on the wavelet used. For the *Haar system*, $n = 1$.

---

[31] Another way would be to perform a 'windowed Fourier transform', which means first looking through a 'window' cut out of the function and then Fourier transform that part. This latter procedure turns out to be considerably more costly in terms of computational operations.

[32] The sum will converge in L^2, and equality will hold almost everywhere, in most cases.

[33] Some authors use the reversed convention with the scale of the variable increasing as the index decreases. With this other convention, the scale of subspaces below will then be increasing instead of decreasing with the index.

[34] A good reference for wavelet theory is the book Ingrid Daubechies, *Ten Lectures on Wavelets*, SIAM, Philadephia, 1992. Daubechies has personally contributed to the development of the theory. See also the introductory book J Bergh, F Ekstedt, M Lindberg, *Wavelets*, Studentlitteratur, Lund, 1999.

In what follows, we describe the arguments for a general orthogonal system of wavelets, which are real-valued and correspond to a finite transform in each step. Then we discuss the algorithms, and we exemplify with the Haar system. Finally, we interprete the analysis in terms of a scale of subspaces of L^2, with pertinent projections and orthogonal complements.

5.1 Wavelets

We start from the function $\phi \in L^1 \bigcap L^2$ with $\int \phi \neq 0$, the **scaling function**, which satisfies a recurrence equation

$$\phi(x) = \sum_{k=0}^{n} c_k 2^{1/2} \phi(2x - k)$$

together with the orthogonality conditions

$$\int \phi(x-m)\phi(x)\, dx = \delta_{m0}$$

The function is thus normalized by $(\int \phi^2)^{1/2} = (\sum_{k=0}^{n} c_k^2)^{1/2} = 1$ (The terms in the right-hand side are orthogonal).

For the Haar system, $\phi = 1_{(0,1)}$, $n = 1$, and $c_0 = c_1 = 2^{-1/2}$.

Note that $0 \leq x \leq n/2$ implies $-k \leq 2x - k \leq n - k$. Thus, if $\phi(x) = 0$ outside the interval $(0, n)$ (see the next exercise), then its values in $(0, n/2)$ are determined by its values at the integers in $(0, n)$. In the Haar system, (ϕ has a jump discontinuity at 0 and at 1), we will put $\phi(0) = 1$ and $\phi(1) = 0$.

A Fourier transformation of the recurrence equation gives

$$\mathcal{F}\phi(s) = \mathcal{F}\phi(s/2)\, 2^{-1/2} \sum_{k=0}^{n} c_k e^{-\pi i k s} := \mathcal{F}\phi(s/2) p(s/2)$$

Note that $p(0) = 1$ follows here. A reiteration gives

$$\mathcal{F}\phi(s) = \mathcal{F}\phi(s/2^N)\, \Pi_{k=1}^{N} p(s/2^k)$$

and, when $N \to \infty$,

$$\mathcal{F}\phi(s) = \mathcal{F}\phi(0)\, \Pi_{k=1}^{\infty} p(s/2^k)$$

In the Haar system, the left-hand side is the function $e^{-\pi i s} \operatorname{sinc} s$, which thus is expressed as an infinite product.

Exercise 5.1.1 *Show that the last equation implies that* $\phi(x) = 0$ *outside* $(0, n)$.

Apparently, the polynomial p or its coefficients c_k contain all information about the function ϕ. In the algorithms, nothing but these coefficients appear.

The ortonormality for the integer translates of ϕ above becomes by Parseval's Formula

$$\begin{aligned}\delta_{m0} &= \int \mathcal{F}\phi(s)\left(e^{-2\pi ims}\mathcal{F}\phi(s)\right)^* ds \\ &= \int_0^1 e^{2\pi ims}\sum_{l=-\infty}^{\infty}|\mathcal{F}\phi(s+l)|^2\, ds\end{aligned}$$

which implies

$$1 \equiv \sum_{l=-\infty}^{\infty}|\mathcal{F}\phi(s+l)|^2 = \sum_{l=-\infty}^{\infty}|\mathcal{F}\phi((s+l)/2)|^2|p((s+l)/2)|^2$$

This yields (consider l even and odd respectively in the sum; p has period 1)

$$|p(s)|^2 + |p(s+1/2)|^2 \equiv 1$$

or

$$\sum_{k=0}^{n} c_{k+2m}c_k = 0 \tag{1}$$

From ϕ, we define the function ψ, **the wavelet**, which will satisfy the two central requirements

$$\int \psi(x)\phi(x-m)\,dx = 0$$

for all integers m, and

$$\psi(x) = \sum_k d_k\, 2^{1/2}\phi(2x-k)$$

These mean that linear combinations of integer translates of ψ are orthogonal to such of ϕ, and that ψ is a linear combination of half-integer translates av ϕ in the halved scale. A normalization is also done here by $(\int \psi^2)^{1/2} = (\sum_k d_k^2)^{1/2} = 1$.

We will now argue that the coefficients d_k are practically determined by the coefficients c_k. Parseval and recurrence give, with $q(s) := 2^{-1/2}\sum_k d_k e^{-2\pi iks}$ and using the same calculations as before (verify!), the condition

$$p(s)^*q(s) + p(s+1/2)^*q(s+1/2) = 0$$

This is fulfilled (essentially only) by

$$q(s) = e^{-2\pi i(s+1/2)}p(s+1/2)^*$$

which implies $q(0) = 0$ by the identity for the polynomial p above and $p(0) = 1$. The condition becomes

$$\sum_k d_k c_{k+2m} = \sum_{k=-n+1}^{1}(-1)^k c_{1-k}c_{k+2m} = 0 \tag{2}$$

(verify!) Consequently, we have

$$\psi(x) = \sum_{k=-n+1}^{1} (-1)^k c_{1-k} 2^{1/2} \phi(2x-k)$$

and $(q(0)=0)$

$$\int \psi(x)\,dx = \mathcal{F}\psi(0) = \mathcal{F}\phi(0)q(0) = 0$$

When $\phi = 1_{(0,1)}$, $\psi = 1_{(0,1/2)} - 1_{(1/2,1)}$ and is called the *Haar function.*

The orthogonality condition $(|k|+|l| \neq 0)$

$$\int \psi(2^{-k}x-l)\psi(x)\,dx = 0$$

is thus fulfilled for $k=0$.

Orthogonality for $k=-1$ follows from

$$\int \psi(2x-l)\psi(x)\,dx = \sum_{m=-n+1}^{1} (-1)^m c_{1-m} \int \psi(2x-l)\phi(2x-m)\,dx = 0$$

In the same way, we get the orthogonality for all $k \neq 0$.

5.2 Fast Wavelet Transform: FWT

The algorithms are founded on the following observation. The equations (1) and (2) may be interpreted as that the matrices L^*L and H^*H below represent projections with orthogonal values (more about this presently). The matrices L and H contain only the coefficients in the polynomial p. [35]

$$[L]_{ij} = c_{j-2i} \qquad\qquad [H]_{ij} = (-1)^j c_{1+2i-j}$$

If the signal to be transformed has length $N=2^K$, we choose $0 \leq i \leq 2^{K-1}-1$ and $0 \leq j \leq 2^K - 1$ in the first step. For the Haar system, if $N = 2^2$ and the signal is the column vector $x = [x_0 \;\; x_1 \;\; x_2 \;\; x_3]^*$ we have

$$\begin{aligned} Lx &= 2^{-1/2}[x_0+x_1 \;\; x_2+x_3]^* \\ L^*Lx &= 1/2[x_0+x_1 \;\; x_0+x_1 \;\; x_2+x_3 \;\; x_2+x_3]^* \\ Hx &= 2^{-1/2}[x_0-x_1 \;\; x_2-x_3]^* \\ H^*Hx &= 1/2[x_0-x_1 \;\; -x_0+x_1 \;\; x_2-x_3 \;\; -x_2+x_3]^* \end{aligned}$$

[35] In the review article by G Strang, Wavelets and dilation equations: a brief introduction, *SIAM Review* **31** (4), 1989, 614-627, the coefficient index in the definition of the matrix L has erroneously been given the wrong sign. The matrix H is similarly wrong. This means that, given the recurrence equation, the matrices in the article belong to the functions $\phi(n-x)$ and $\psi(n-x)$.

The notation L and H are chosen to indicate *low-pass* and *high-pass* filter respectively; the reason is explained below.

The following matrix conditions are equivalent to the polynomial ones for p and q (which in turn are equivalent to (1) and (2)).

$$HL^* = 0 \quad (\& \quad LH^* = 0) \tag{3}$$
$$LL^* = E \quad \& \quad HH^* = E \tag{4}$$
$$L^*L + H^*H \quad = \quad E \tag{5}$$

The wavelet anaysis of the signal x means, in step 1, to calculate Lx and Hx, which both will be half as long as x. The last step of the reconstruction will be to calculate L^*Lx and H^*Hx from Lx and Hx, and sum them: $L^*Lx + H^*Hx = x$. If the signal length is $N = 2^K$, the analysis will be a reiteration (at most) K times of step 1 on the result of operating with L in the previous step; for the reconstruction the corresponding procedure is applied.

Exercise 5.2.1 *Do the entire analysis above for $N = 2^2$ and compare step by step to the FFT.*

The vector $x - L^*Lx = H^*Hx$ is the projection on a 'high-frequency' component of x, while $x - H^*Hx = L^*Lx$ is the projection on a 'low-frequency' component. The components are orthogonal by (3):

$$H^*Hx(L^*Lx)^* = H^*Hxx^*L^*L = H^*HL^*Lxx^* = 0$$

In addition, we have from (4)

$$(H^*H)^2 = H^*HH^*H = H^*H \;\; \& \;\; (L^*L)^2 = L^*LL^*L = L^*L$$

that is, H^*H and L^*L are projections.

The condition (5) therefore means that we can divide any vector into orthogonal components: one with the high frequencies and one with the low.

A natural question, which we now turn to, is:

> **What is the connection between the results of the steps in the algorithm and the sample values and the wavelet coefficients of the continuous-time signal?**

Write the low-pass filtered continuous-time signal $x(t)$, with $\mathcal{F}x(s) = 0$ for $|s| \geq 1/2$, and where the sample values $x(l)$ are approximated to 0 outside the integers $0, 1, 2, \ldots, 2^K - 1$, as (the index a as in *approximation*)

$$x_a = \text{sinc} * \sum_{l=0}^{2^K-1} x(l)\delta_l = \sum_{l=0}^{2^K-1} x(l)\text{sinc}(\cdot - l)$$

The starting point of the analysis is (the index w as in *wavelet*)

$$x_w = \phi * \sum_{l=0}^{2^K-1} x(l)\delta_l = \sum_{l=0}^{2^K-1} x(l)\phi(\,\cdot\, - l))$$

After a Fourier transformation, we have

$$\begin{aligned}
\mathcal{F}x_a(s) &= 1_{(-1/2,1/2)}(s) \sum_{l=0}^{2^K-1} x(l)e^{-2\pi i s l} \\
\mathcal{F}x_w(s) &= \mathcal{F}\phi(s) \sum_{l=0}^{2^K-1} x(l)e^{-2\pi i s l}
\end{aligned}$$

The function x_w will apparently contain the same information as the function x_a precisely when the condition $\mathcal{F}\phi(s) \neq 0 \;\Leftrightarrow\; 1_{(-1/2,1/2)}(s) \neq 0$ is fulfilled.

REMARK 1 The wavelet analysis is thus not performed directly on the low-pass filtered signal, but after a filtering which adds alias effects if $\mathcal{F}\phi(s) \neq 0$ outside $(-1/2, 1/2)$, most prominently in the smallest scales. High (low) frequencies will then seem to contain more (less) energy than their part of the energy spectrum of the signal. If $\mathcal{F}\phi(s)$ decays slowly to 0 outside $(-1/2, 1/2)$, this influence will be clear. To avoid such effects in the analysis result, a pre-filtering can be performed, or perhaps choose another ϕ to start with.

Exercise 5.2.2 *Show that, in $\mathcal{S}'$,*

$$\sum_{l=0}^{2^K-1} x_l\, n\phi(n(\,\cdot\, - l)) \to \mathcal{F}\phi(0) \sum_{l=0}^{2^K-1} x_l\delta_l$$

How does a change of $\phi(x)$ to $n^{1/2}\phi(nx)$ affect the recurrence equation?

Consequently, we now assume that $(x_l := x(l))$

$$x_w = \sum_{l=0}^{2^K-1} x_l\phi(\,\cdot\, - l))$$

represents (after filtering) the usual approximation to finitely many sample values of the given low-pass filtered continuous-time signal x.

To simplify the notation, we will in the sequel use the scalar product in L^2

$$(f, g) := \int f(x)g(x)^*\, dx$$

and, *e.g.*,

$$\phi_{k,l}(x) := 2^{-k/2}\phi(2^{-k}x - l)$$

Given the coefficients x_j, $j = 0, 1, 2, ..., 2^K - 1$, the relation of which to the signal x was just discussed, we have x_w as the input for the analysis

$$x_w = \sum_{j=0}^{2^K-1} x_j \phi_{0,j}$$

The wavelet coefficients $a_{1,i}$ are then, using definitions and orthogonality, given by

$$\begin{aligned}
a_{1,i} &= (x_w, \psi_{1,i}) = \sum_{j=0}^{2^K-1} x_j (\phi_{0,j}, \psi_{1,i}) \\
&= \sum_{j=0}^{2^K-1} x_j \sum_{k=-n+1}^{1} (-1)^k c_{1-k} (\phi_{0,j}, \phi_{0,2i+k}) \\
&= \sum_{j=0}^{2^K-1} x_j (-1)^j c_{1+2i-j}
\end{aligned}$$

Here we have now seen the matrix H in action. The matrix L acts in a corresponding way:

$$\begin{aligned}
b_{1,i} &= (x_w, \phi_{1,i}) = \sum_{j=0}^{2^K-1} x_j (\phi_{0,j}, \phi_{1,i}) \\
&= \sum_{j=0}^{2^K-1} x_j \sum_{k=0}^{n} c_k (\phi_{0,j}, \phi_{0,2i+k}) \\
&= \sum_{j=0}^{2^K-1} x_j c_{j-2i}
\end{aligned}$$

The terms *high-pass* and *low-pass* filter is motivated by $\mathcal{F}\psi(0) = 0$ and $\mathcal{F}\phi(0) \neq 0$, respectively (both are more or less localized around 0) in the equality

$$\begin{aligned}
\mathcal{F}x_w &= \sum_{l=0}^{2^{K-1}-1} a_{1,l} \mathcal{F}\psi_{1,l} + \sum_{l=0}^{2^{K-1}-1} b_{1,l} \mathcal{F}\phi_{1,l} \\
&= 2\mathcal{F}\psi(2s) \sum_{l=0}^{2^{K-1}-1} a_{1,l} e^{-4\pi i l s} + 2\mathcal{F}\phi(2s) \sum_{l=0}^{2^{K-1}-1} b_{1,l} e^{-4\pi i l s}
\end{aligned}$$

In the Haar system, $\mathcal{F}\phi(s) = e^{-\pi i s} \operatorname{sinc} s$.

Exercise 5.2.3 *Show that, in the Haar system,*

$$\mathcal{F}\psi(s) = i\, e^{-\pi i s} \sin(\pi s/2) \operatorname{sinc}(s/2)$$

5.3 A Scale of Subspaces

Let ϕ and ψ be as before, and put V_0 as the set of all finite linear combinations of integer translates of ϕ, $\phi(x-k)$, together with their limits in L^2.

Let V_1 be the corresponding with $2^{-1/2}\phi(2^{-1}x-k)$: doubled scale – thus the index 1 (>0). The recurrence equation gives $V_1 \subset V_0$. By definition, the set W_1 of all finite linear combinations of the integer translates $2^{-1/2}\psi(2^{-1}x-k)$, and their limits in L^2, is a subset of V_0. From the equalities (1) and (2), we infer

$$V_0 = W_1 \bigoplus V_1 \quad \text{and} \quad W_1 \perp V_1$$

With V_k and W_k as the set of finite linear combinations of the integer translates $2^{-k/2}\phi(2^{-k}x-l)$ and $2^{-k/2}\psi(2^{-k}x-l)$, respectively, and their limits in L^2, we have a scale of subspaces in L^2,

$$\{0\} \subset \ ... \ \subset V_{k+1} \subset V_k \subset V_{k-1} \subset \ ... \ \subset L^2$$

for all integers k, where

$$V_{k-1} = W_k \bigoplus V_k \quad \text{and} \quad W_k \perp V_k$$

Furthermore, it can be shown that

$$\bigcup_k V_k = L^2 \quad \text{and} \quad \bigcap_k V_k = \{0\}$$

where also limits in L^2 are counted as belonging to the union ('the closed hull is L^2').

If we let P_k and Q_k denote the orthogonal projections of V_{k-1} on V_k and W_k, respectively, then $P_k + Q_k$ becomes the identity mapping on V_{k-1}.

In the algorithms, the equality

$$V_0 = W_1 \bigoplus W_2 \bigoplus \ \ ... \ \bigoplus W_K \bigoplus V_K$$

was used together with the projections P_k and Q_k. The signal length is 2^K, and it is supposed to belong to V_0. The analysis algorithm means successive projections on the orthogonal subspaces. The factor H in the projection $Q_k = H^*H$ gives the wavelet coefficients as coefficents in the linear combination of $\psi(2^{-k}x-l)$, that is, in the scale 2^k.

A Appendix

Here we give a proof of the Structure Theorem 1.2 for tempered distributions. A characterization concludes the appendix.

Theorem A.1 (The Structure Theorem 1.2) *Let $T \in \mathcal{S}'$. Then continuous functions f_j, $j = 1, 2, \ldots$, and non-negative integers β_j, exist such that (in $\mathcal{S}'$)*

$$T = \sum_{j=-\infty}^{\infty} D^{\beta_j} f_j$$

The somewhat technical proof is based on Lemma A.1.1 – 3, and on a result (which we do not prove here) from Integration Theory. (None of these results will be used in the rest of the course.)

Lemma A.1.1 *Let $T \in \mathcal{S}'$ and $\varphi \in \mathcal{S}$ with $\varphi(x) = 0$ for $x \notin (a, b)$, which is a finite interval.*

Then a non-negative integer β and a constant C exist so that (for these φ)

$$| < T, \varphi > | \leq C \sup_x |D^\beta \varphi(x)|$$

Proof of Lemma A.1.1: Assume the contrary, that is, $\beta_n \to \infty$ and φ_n, $n = 1, 2, \ldots$, exist such that

$$\sup_x |D^{\beta_n} \varphi_n(x)| \longrightarrow 0$$

but

$$| < T, \varphi_n > | \geq 1$$

Now $\varphi_n \to 0$ in $\mathcal{S}$ follows, since, given α, β, take n such that $\beta_n \geq \beta$, and we get

$$\begin{aligned} \sup_x (1 + |x|)^\alpha |D^\beta \varphi_n(x)| &\leq C_{a,b} \sup_x |D^\beta \varphi_n(x)| \\ &\leq C_{a,b} \sup_x |D^{\beta_n} \varphi_n(x)| \longrightarrow 0 \end{aligned}$$

where the last inequality derives from

$$|\varphi_n(x)| = | \int_a^x D\varphi_n(y)\, dy| \leq (b - a) \sup_x |D\varphi_n(x)|$$

We have now a contradiction: $\varphi_n \to 0$ in $\mathcal{S}$ and $| < T, \varphi_n > | \geq 1$. This proves the lemma.

□

The smallest admissible β in Lemma A.1.1 is called the *order* of T in (a, b). When the order is 0, T is called a *(tempered) measure* in (a, b).

Lemma A.1.2 *Assume that DT has order $\beta \geq 1$ in (a, b). Then T has order $\beta - 1$.*

Proof of Lemma A.1.2: Let DT have order $\beta \geq 1$ in (a, b), and let φ be as in Lemma A.1.1. We show that T then has order at most $\beta - 1$.

We have

$$| < DT, \varphi > | \leq C \sup_x |D^\beta \varphi(x)|$$

and thus

$$| < T, D\varphi > | \leq C \sup_x |D^\beta \varphi(x)|$$

which is the desired inequality, though only in a subspace of those functions which are derivatives. Fix now φ_0 as in Lemma A.1.1 with $\int \varphi_0(x)\, dx = 1$. Taking ψ as in Lemma A.1.1, we have [36]

$$\psi(x) = D\{\int_a^x \psi(y)\, dy - \int_a^x \varphi_0(y)\, dy \int_a^b \psi(y)\, dy\} + \varphi_0(x) \int_a^b \psi(y)\, dy$$

which yields (verify that the expression in curly brackets has the desired properties!)

$$\begin{aligned}
| < T, \psi > | &\leq C\{\sup_x |D^\beta (\int_a^x \psi(y)\, dy - \int_a^x \varphi_0(y)\, dy \int_a^b \psi(y)\, dy| \\
&\quad + | < T, \varphi_0 > \int_a^b \psi(y)\, dy|\} \\
&\leq C\{\sup_x (|D^{\beta-1}\psi(x)| + |D^{\beta-1}\varphi_0(x)|| \int_a^b \psi(y)\, dy)|) \\
&\quad + | \int_a^b \psi(y)\, dy)|\} \\
&\leq C \sup_x |D^{\beta-1}\psi(x)|
\end{aligned}$$

where the last inequality is obtained as in the proof of Lemma A.1.1. This shows that the order of T is at most $\beta - 1$. (Verify 'at most'!) The proof is done.

□

We will now prove the existence of a *primitive distribution.*

Lemma A.1.3 *Let $T \in \mathcal{S}'$. Then $S \in \mathcal{S}'$ exists with $DS = T$.*

Proof of Lemma A.1.3: If S existed, we would have ($\varphi \in \mathcal{S}$)

$$< T, \varphi > = < DS, \varphi > = - < S, D\varphi >$$

[36] Compare the proof of Fourier's Inversion Formula.

This equation defines S on the subspace of derivatives in $\mathcal{S}$. Taking φ_0 as in the proof of Lemma A.1.2, we write again ($\psi \in \mathcal{S}$)

$$\psi(x) = D\{\int_{-\infty}^{x} \psi(y)\,dy - \int_{-\infty}^{x} \varphi_0(y)\,dy \int_{-\infty}^{\infty} \psi(y)\,dy\} + \varphi_0(x) \int_{-\infty}^{\infty} \psi(y)\,dy$$

We define $S(\varphi_0) = k$ and now have S defined on the whole $\mathcal{S}$. It remains to show that $\psi_n \to 0$ in $\mathcal{S}$ entails $S(\psi_n) \to 0$, which is done as in the proof of Lemma A.1.1. (Verify that the object to be differentiated is in $\mathcal{S}$, and the last statement!)

□

Proof of Theorem 1.2: To enable the use of Lemma A.1.1, where the interval is finite, we partition T into a sum.

Take $\Psi \in \mathcal{S}$ such that $\Psi(x) > 0$ when $x \in (-1, 1)$, and $\Psi(x) = 0$ otherwise. (Cf. Example 1.1.2.) Put $\psi_n(x) = \Psi(x-n)/\sum_\nu \Psi(x-\nu)$ with n integer. This constitutes a *partition of unity*, that is, $\sum_n \psi_n = 1$ and $\psi_n \in \mathcal{S}$ (verify!).

By Proposition 1.2.1, we may write $T = \sum_n \psi_n T$ in $\mathcal{S}'$ (verify!). Each term gives, for $\varphi \in \mathcal{S}$ with $\varphi(x) = 0$ outside $(n-1, n+1)$, invoking Lemma A.1.1,

$$\begin{aligned} |< \psi_n T, \varphi >| &= |< T, \psi_n \varphi >| \\ &\le C \sup_x |D^{\beta_n}\{\psi_n(x)\varphi(x)\}| \\ &\le C \sup_x |D^{\beta_n}\varphi(x)| \end{aligned}$$

that is, $\psi_n T$ has order β_n in $(n-1, n+1)$. By Lemma A.1.2 and Lemma A.1.3, there is a measure S_n with $D^{\beta_n} S_n = \psi_n T$. Now we cite, without proof, a result of which the proof requires knowledge about the Lebesgue integral: *Every measure is (in $\mathcal{S}'$) the second derivative of a continuous function* This results provides a continuous function f_n with $D^{\beta_n+2} f_n = \psi_n T$ i $\mathcal{S}'$. Consequently, we have (with new β_n two units bigger)

$$T = \sum_n D^{\beta_n} f_n$$

and the proof is complete.

□

A characterization of tempered distributions is also obtained using the same technique.

Proposition A.1.1 *Let $T : \mathcal{S} \to \boldsymbol{C}$ be linear. Then T is a tempered distribution if, and only if, a constant C and integers α, β exist, such that ($\varphi \in \mathcal{S}$)*

$$|T(\varphi)| \le C \sup_x \left[(1 + |x|)^\alpha \sum_{k \le \beta} |D^k \varphi(x)|\right]$$

Proof: The 'if' part follows immediately from the definition of tempered distribution. The 'only if' part follows by an obvious modification of the proof of Lemma A.1.1. (Verify!)

□

Index