Chapter 4

Introduction to the Fourier transform

In this chapter we introduce the Fourier transform and review some of its basic properties. The Fourier transform is the "swiss army knife" of mathematical analysis; it is a powerful general purpose tool with many useful special features. In particular the theory of the Fourier transform is largely independent of the dimension: the theory of the Fourier transform for functions of one variable is formally the same as the theory for functions of 2, 3 or n variables. This is in marked contrast to the Radon, or X-ray transforms. For simplicity we begin with a discussion of the basic concepts for functions of a single variable, though in some definitions, where there is **no** additional difficulty, we treat the general case from the outset.

4.1 The complex exponential function.

See: 2.2, A.4.3 .

The building block for the Fourier transform is the complex exponential function, e^{ix}. The basic facts about the exponential function can be found in section A.4.3. Recall that the polar coordinates (r, θ) correspond to the point with rectangular coordinates $(r\cos\theta, r\sin\theta)$. As a complex number this is

$$r(\cos\theta + i\sin\theta) = re^{i\theta}.$$

Multiplication of complex numbers is very easy using the polar representation. If $z = re^{i\theta}$ and $w = \rho e^{i\phi}$ then

$$zw = re^{i\theta}\rho e^{i\phi} = r\rho e^{i(\theta+\phi)}.$$

A positive number r has a real logarithm, $s = \log r$, so that a complex number can also be expressed in the form

$$z = e^{s+i\theta}.$$

The logarithm of z is therefore defined to be the complex number

$$\log z = s + i\theta = \log|z| + i\tan^{-1}\left(\frac{\operatorname{Im} z}{\operatorname{Re} z}\right).$$

As $\exp(2\pi i) = 1$, the imaginary part of the $\log z$ is only determined up to integer multiplies of 2π.

Using the complex exponential we build a family of functions, $\{e^{ix\xi} : \xi \in \mathbb{R}\}$. Sometimes we think of x as the variable and ξ as a parameter and sometimes their roles are interchanged. Thinking of ξ as a parameter we see that $e^{ix\xi}$ is a $\frac{2\pi}{\xi}$-periodic function, that is

$$\exp(i(x + \frac{2\pi}{\xi})\xi) = \exp(ix\xi).$$

In physical applications $e^{ix\xi}$ describes an oscillatory state with *frequency* $\frac{\xi}{2\pi}$ and *wave length* $\frac{2\pi}{\xi}$. The goal of Fourier analysis is to represent "arbitrary" functions as linear combinations of these oscillatory states. Using (A.57) we easily derive the fact that

$$\partial_x e^{ix\xi} = i\xi e^{ix\xi}. \tag{4.1}$$

Loosely speaking this formula says that $e^{ix\xi}$ is an *eigenvector* with *eigenvalue* $i\xi$ for the linear operator ∂_x. In quantum mechanics $e^{ix\xi}$ is a state with *momentum* ξ and *energy* $\frac{|\xi|^2}{2}$.

Exercises

Exercise 4.1.1. If a is a real number then it is a consequence of the Fundamental Theorem of Calculus that

$$\int_0^x e^{ay} dy = \frac{e^{ax} - 1}{a} \tag{4.2}$$

Use the power series for the exponential to prove that this formula remains correct, even if a is a complex number.

Exercise 4.1.2. Use the power series for the exponential to prove that (4.1) continues to hold for ξ any complex number.

Exercise 4.1.3. Use the differential equation satisfied by e^x to show that $e^x e^{-x} = 1$. Hint: Use the uniqueness theorem for solutions of ODEs.

Exercise 4.1.4. If $\operatorname{Re} a < 0$ then the improper integral is absolutely convergent:

$$\int_0^\infty e^{ax} dx = \frac{-1}{a}.$$

Using the triangle inequality (not the explicit formula) show that

$$\left| \int_0^\infty e^{ax} dx \right| \leq \frac{1}{|\operatorname{Re} a|}.$$

Exercise 4.1.5. Which complex numbers have purely imaginary logarithms?

4.2 The Fourier transform for functions of a single variable

We now turn our attention to the Fourier transform for functions of a single real variable. As the complex exponential itself assumes complex values, it is natural to consider complex valued functions from the outset. The theory for functions of several variables is quite similar and is treated later in the chapter.

4.2.1 Absolutely integrable functions

See: 2.2.2, A.5.1.

Let f be a function defined on $\mathbb{R}^n$, recall that f is *absolutely integrable* if

$$\|f\|_1 = \int_{\mathbb{R}^n} |f(\mathbf{x})| d\mathbf{x} < \infty.$$

The set of such functions is a vector space. It would be natural to use $\| \cdot \|_1$ to define a norm on this vector space, but there is a small difficulty: If f is supported on a set of measure zero then $\|f\|_1 = 0$. In other words there are non-zero, absolutely integrable functions with "norm" zero. As real measurements are usually expressed as integrals, two functions which differ on a set of measure zero cannot be distinguished by any practical measurement. For example the functions $\chi_{[0,1]}$ and $\chi_{[0,1)}$ are indistiguishable from the point of view of measurements and of course

$$\|\chi_{[0,1]} - \chi_{[0,1)}\|_1 = 0.$$

The way to circumvent this technical problem is to *declare* that two absolutely integrable functions, f_1 and f_2 are the *same* whenever $f_1 - f_2$ is supported on a set of measure zero. In other words, we identify two states which cannot be distinguished by any realistic measurment. This defines a equivalence relation on the set of integrable functions. The normed vector space $L^1(\mathbb{R}^n)$ is defined to be the set of absolutely integrable functions modulo this equivalence relation with norm defined by $\| \cdot \|_1$. This is a complete, normed linear space.

This issue arises whenever an integral is used to define a norm. Students unfamiliar with this concept need not worry: As it plays very little role in imaging (or mathematics, for that matter) we will usually be sloppy and ignore this point, acting as if the elements of $L^1(\mathbb{R}^n)$ and similar spaces are ordinary functions.

4.2.2 The Fourier transform for integrable functions

The natural domain for the Fourier transform is the space of absolutely integrable functions.

Definition 4.2.1. The Fourier transform of an absolutely integrable function f, defined on $\mathbb{R}$ is the function $\hat{f}$ defined on $\mathbb{R}$ by the integral

$$\hat{f}(\xi) = \int_{-\infty}^{\infty} f(x) e^{-ix\xi} dx. \tag{4.3}$$

The utility of the Fourier transform stems from the fact that f can be "reconstructed" from $\hat{f}$. A result that suffices for most of our applications is the following:

Theorem 4.2.1 (Fourier inversion formula). *Suppose that f is an absolutely integrable function such that $\hat{f}$ is also absolutely integrable, then*

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{f}(\xi)e^{ix\xi}d\xi. \tag{4.4}$$

Remark 4.2.1. Formula (4.4) is called the *Fourier inversion formula.* It is *the* prototype of all reconstruction formulæ used in medical imaging.

Proof. We give a proof of the inversion formula under the additional assumption that f is continuous, this assumption is removed in section 5.1. We need to show that

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{f}(\xi)e^{i\xi x}d\xi.$$

Because $\hat{f}$ is in $L^1(\mathbb{R})$ it is not difficult to show that

$$\begin{aligned}\frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{f}(\xi)e^{i\xi x}d\xi &= \lim_{\epsilon\to 0^+}\frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{f}(\xi)e^{-\epsilon\xi^2}e^{i\xi x}d\xi \\ &= \lim_{\epsilon\to 0^+}\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(y)e^{-\epsilon\xi^2}e^{i\xi(x-y)}dyd\xi.\end{aligned} \tag{4.5}$$

Interchange the integrations in the last formula we use example 4.2.4 to compute the Fourier transform of the Gaussian to get

$$\begin{aligned}\lim_{\epsilon\to 0^+}\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(y)e^{-\epsilon\xi^2}e^{i\xi(x-y)}dyd\xi &= \lim_{\epsilon\to 0^+}\frac{1}{2\sqrt{\epsilon\pi}}\int_{-\infty}^{\infty} f(y)e^{-\frac{(x-y)^2}{4\epsilon}}dy \\ &= \lim_{\epsilon\to 0^+}\frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty} f(x-2\sqrt{\epsilon}t)e^{-t^2}dt.\end{aligned} \tag{4.6}$$

As f is continuous and integrable it follows that the limit in the last line is

$$\frac{f(x)}{\sqrt{\pi}}\int_{-\infty}^{\infty} e^{-t^2}dt.$$

The proof is completed by observing that this integral equals $\sqrt{\pi}$. □

Remark 4.2.2. The Fourier transform and its inverse are integral transforms which are frequently thought of as mappings. In this context it is customary to use the notation:

$$\begin{aligned}\mathcal{F}(f) &= \int_{-\infty}^{\infty} f(x)e^{-ix\xi}dx, \\ \mathcal{F}^{-1}(f) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathcal{F}(f)(\xi)e^{ix\xi}d\xi.\end{aligned} \tag{4.7}$$

Observe that the operation performed to recover f from $\hat{f}$ is almost the same as the operation performed to obtain $\hat{f}$ from f. Indeed if $f_r(x) \stackrel{d}{=} f(-x)$ then

$$\mathcal{F}^{-1}(f) = \frac{1}{2\pi}\mathcal{F}(f_r). \tag{4.8}$$

This symmetry accounts for many of the Fourier transform's remarkable properties. As the following example shows, the assumption that f is in $L^1(\mathbb{R})$ does not imply that $\hat{f}$ is as well.

Example 4.2.1. Define the function

$$r_1(x) = \begin{cases} 1 \text{ for } -1 < x < 1, \\ 0 \text{ for } 1 < |x|. \end{cases} \tag{4.9}$$

The Fourier transform of r_1 is

$$\widehat{r_1}(\xi) = \int_{-1}^{1} e^{-\xi x} dx = \frac{1}{-i\xi} e^{-\xi x}\Big|_{-1}^{1} = \frac{2\sin\xi}{\xi},$$

and

$$\int_{-\infty}^{\infty} |\widehat{r_1}(\xi)| d\xi = 2 \int_{-\infty}^{\infty} \frac{|\sin\xi|}{|\xi|}$$

diverges. So while r_1 is absolutely integrable its Fourier transform $\hat{r}_1$ is not. As the Fourier transform of r_1 is such an important function in image processing, we define

$$\operatorname{sinc}(x) \stackrel{d}{=} \frac{\sin(x)}{x}.$$

Example 4.2.2. Recall that $\chi_{[a,b)}(x)$ equals 1 for $a \leq x < b$ and zero otherwise. Its Fourier transform is given by

$$\hat{\chi}_{[a,b)}(\xi) = \frac{e^{-ib\xi} - e^{-ia\xi}}{i\xi}. \tag{4.10}$$

Example 4.2.3. A family of functions arising in magentic resonance imaging are those of the form

$$f(x) = \chi_{[0,\infty)}(x) e^{i\alpha x} e^{-\beta x}, \ \alpha \in \mathbb{R} \text{ and } \beta > 0.$$

By simply computing the integral we find that

$$\hat{f}(\xi) = \frac{1}{\beta + i(\xi - \alpha)}$$

Using the fact that $e^{i\alpha x} = \cos(\alpha x) + i\sin(\alpha x)$ it is not difficult to show that

$$\begin{gathered} \mathcal{F}(\cos(\alpha x) e^{-\beta x} \chi_{[0,\infty)}(x)) = \frac{\beta + i\xi}{\beta^2 + \alpha^2 - \xi^2 + 2i\xi\beta} \\ \text{and} \\ \mathcal{F}(\sin(\alpha x) e^{-\beta x} \chi_{[0,\infty)}(x)) = \frac{\alpha}{\beta^2 + \alpha^2 - \xi^2 + 2i\xi\beta}. \end{gathered} \tag{4.11}$$

Example 4.2.4. The "Gaussian," e^{-x^2} is a function of considerable importance in image processing and mathematics. Its Fourier transform was already used in the proof of the inversion formula. For later reference we record its Fourier transform:

$$\begin{aligned}\mathcal{F}(e^{-x^2})(\xi) &= \int\limits_{-\infty}^{\infty} e^{-x^2} e^{-i\xi x} dx \\ &= \sqrt{\pi} e^{-\frac{\xi^2}{4}},\end{aligned} \tag{4.12}$$

or more generally

$$\mathcal{F}(e^{-ax^2})(\xi) = \sqrt{\frac{\pi}{a}} e^{-\frac{\xi^2}{4a}}. \tag{4.13}$$

This is derived in section 4.2.3. Note that $e^{-\frac{x^2}{2}}$ is an eigenvector of the Fourier transform, with eigenvalue $\sqrt{2\pi}$.

4.2.3 Appendix: The Fourier transform of a Gaussian*

For completeness we include a derivation of the Fourier transform of the Gaussian e^{-x^2}. It uses the Cauchy integral formula for analytic functions of a complex variable. The Fourier transform is given by

$$\begin{aligned}\mathcal{F}(e^{-x^2})(\xi) &= \int\limits_{-\infty}^{\infty} e^{-(x^2+ix\xi)} dx \\ &= e^{-\frac{\xi^2}{4}} \int\limits_{-\infty}^{\infty} e^{-(x+i\xi/2)^2} dx.\end{aligned} \tag{4.14}$$

The second integral is the complex contour integral of the analytic function e^{-z^2} along the contour $\operatorname{Im} z = \xi/2$. Because e^{-z^2} decays rapidly to zero as $|\operatorname{Re} z|$ tends to infinity, Cauchy's theorem implies that the contour can be shifted to the real axis without changing the value of the integral, that is

$$\int\limits_{-\infty}^{\infty} e^{-(x+i\xi/2)^2} dx = \int\limits_{-\infty}^{\infty} e^{-x^2} dx. \tag{4.15}$$

To compute the last integral observe that

$$\begin{aligned}\left[\int\limits_{-\infty}^{\infty} e^{-x^2} dx\right]^2 &= \int\limits_{-\infty}^{\infty} e^{-x^2} dx \int\limits_{-\infty}^{\infty} e^{-y^2} dy \\ &= \int\limits_{0}^{2\pi} \int\limits_{0}^{\infty} e^{-r^2} r dr d\theta \\ &= 2\pi \int\limits_{0}^{\infty} e^{-s} \frac{ds}{2} \\ &= \pi.\end{aligned} \tag{4.16}$$

Polar coordinates are used in the second line; in the second to last line we let $s = r^2$. Combining these formulæ gives

$$\mathcal{F}(e^{-x^2}) = \sqrt{\pi} e^{-\frac{\xi^2}{4}}.$$

Exercises

Exercise 4.2.1. Show that if f is a continuous, absolutely integrable function then

$$\lim_{\epsilon \to 0^+} \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} f(x - 2\sqrt{\epsilon} t) e^{-t^2} dt = f(x).$$

Exercise 4.2.2. Suppose that f is absolutely integrable, show that $\hat{f}$ is a bounded, continuous function.

Exercise 4.2.3. Prove the identity (4.8).

Exercise 4.2.4. Prove formula (4.10). Show that for any numbers $a < b$ there is a constant M so that

$$|\hat{\chi}_{[a,b)}(\xi)| \leq \frac{M}{1 + |\xi|}.$$

Exercise 4.2.5. Prove the formulæ in (4.11) and show that

$$\mathcal{F}(e^{-\beta|x|} e^{i\alpha x}) = \frac{2\beta}{\beta^2 + (\xi - \alpha)^2}.$$

Exercise 4.2.6. Derive the formula for $\mathcal{F}(e^{-ax^2})$ from the formula for $\mathcal{F}(e^{-x^2})$.

Exercise 4.2.7. Show that for any $k \in \mathbb{N} \cup \{0\}$ the function $h_k = (\partial_x - x)^k e^{-\frac{x^2}{2}}$ is an eigenfunction of the Fourier transform. That is $\mathcal{F}(h_k) = \lambda_k h_k$. Find λ_k. Hint: Integrate by parts and use induction. Find formulæ for h_1, h_2, h_3.

Exercise 4.2.8. * Give a detailed justification for (4.15).

4.2.4 Regularity and decay

See: A.5.1, A.7.1.

It is a general principle that the *regularity* properties of f are reflected in the *decay* properties of its Fourier transform $\hat{f}$ and similarly the regularity of the Fourier transform is a reflection of the decay properties of f. Without any regularity, beyond absolute integrability we have the fundamental result:

Lemma 4.2.1 (The Riemann-Lebesgue Lemma). *If f is an absolutely integrable function then its Fourier transform $\hat{f}$ is a continuous function which goes to zero at infinity. That is, for $\eta \in \mathbb{R}$,*

$$\lim_{\xi \to \eta} \hat{f}(\xi) = \hat{f}(\eta) \text{ and } \lim_{\xi \to \pm\infty} \hat{f}(\xi) = 0. \tag{4.17}$$

Proof. The second statement is a consequence of the basic approximation theorem for L^1-functions, Theorem A.7.2. According to this theorem, given $\epsilon > 0$ there is a step function F, given by

$$F(x) = \sum_{j=1}^{N} c_j \chi_{[a_j,b_j)}(x).$$

so that

$$\int_{-\infty}^{\infty} |f(x) - F(x)| < \epsilon.$$

Estimating the difference of their Fourier transforms gives

$$\begin{aligned} |\hat{F}(\xi) - \hat{f}(\xi)| &= \left| \int_{-\infty}^{\infty} (F(x) - f(x)) e^{-ix\xi} dx \right| \\ &\leq \int_{-\infty}^{\infty} |F(x) - f(x)| dx \\ &\leq \epsilon. \end{aligned} \tag{4.18}$$

Since ϵ is an arbitrary positive number, it therefore suffices to show that $\lim_{|\xi|\to\infty} \hat{F}(\xi) = 0$. The Fourier transform of F is

$$\begin{aligned} \hat{F}(\xi) &= \sum_{j=1}^{N} c_j \hat{\chi}_{[a_j,b_j)}(\xi) \\ &= \sum_{j=1}^{N} c_j \frac{e^{-ib_j\xi} - e^{-ia_j\xi}}{i\xi}. \end{aligned} \tag{4.19}$$

The second line shows that there is a constant C so that

$$|\hat{F}(\xi)| \leq \frac{C}{1 + |\xi|}.$$

The continuity of $\hat{f}(\xi)$ is left as an exercise. □

To go beyond (4.17) we need to introduce quantitative measures of regularity and decay. A simple way to measure regularity is through differentiation: the more derivatives a function has, the more regular it is.

Definition 4.2.2. For $j \in \mathbb{N} \cup \{0\}$, the set of functions on $\mathbb{R}$ with j continuous derivatives is denoted by $\mathcal{C}^j(\mathbb{R})$. The set of infinitely differentiable functions is denoted by $\mathcal{C}^\infty(\mathbb{R})$.

Since the Fourier transform involves integration over the whole real line it is important to assume that these derivatives are also integrable. To quantify rates of decay we compare a function f to a simpler function such as a power of $\|\mathbf{x}\|$.

Definition 4.2.3. A function f, defined on $\mathbb{R}^n$, decays like $\|\mathbf{x}\|^{-\alpha}$ if there are constants C and R so that

$$|f(\mathbf{x})| \leq \frac{C}{\|\mathbf{x}\|^\alpha} \text{ for } \|\mathbf{x}\| > R.$$

This is sometimes denoted by "$f = O(\|\mathbf{x}\|^{-\alpha})$ as $\|\mathbf{x}\|$ tends to infinity."

The elementary formula for integration by parts is often useful in Fourier analysis. Let f and g be differentiable functions on the interval $[a, b]$ then

$$\int_a^b f'(x)g(x)dx = f(x)g(x)\Big|_{x=a}^{x=b} - \int_a^b f(x)g'(x)dx. \tag{4.20}$$

We need an extension of this formula with $a = -\infty$ and $b = \infty$. For our purposes it suffices to assume that fg, $f'g$ and fg' are absolutely integrable, the integration by parts formula then becomes

$$\int_{-\infty}^{\infty} f'(x)g(x)dx = -\int_{-\infty}^{\infty} f(x)g'(x)dx. \tag{4.21}$$

This formula follows by taking letting a and b tend to infinity in (4.20). That the integrals converge is an immediate consequence of the assumption that $f'g$ and fg' are absolutely integrable. The assumption that fg is also absolutely integrable implies the existence of sequences $< a_n >$ and $< b_n >$ so that

$$\begin{aligned} \lim_{n\to\infty} a_n = -\infty \text{ and } \lim_{n\to\infty} b_n = \infty, \\ \lim_{n\to\infty} fg(a_n) = 0 \text{ and } \lim_{n\to\infty} fg(b_n) = 0. \end{aligned} \tag{4.22}$$

Taking the limits in (4.20) along these sequences gives (4.21).

Suppose that f is an absolutely integrable function with an absolutely integrable first derivative, that is

$$\int_{-\infty}^{\infty} [|f(x)| + |f'(x)|]dx < \infty.$$

Provided $\xi \neq 0$ we can use (4.21) to obtain a formula for $\hat{f}(\xi)$:

$$\begin{aligned} \hat{f}(\xi) &= \int_{-\infty}^{\infty} f(x)e^{-ix\xi}dx \\ &= \int_{-\infty}^{\infty} f'(x)\frac{e^{-ix\xi}}{i\xi}dx. \end{aligned} \tag{4.23}$$

That is

$$\hat{f}(\xi) = \frac{\widehat{f'}(\xi)}{i\xi}.$$

Because f' is absolutely integrable the Riemann-Lebesgue lemma implies that $\widehat{f'}$ tends to zero as $|\xi|$ tends to ∞. Combining our formula for $\hat{f}$ with this observation we see that $\hat{f}$ goes to zero more rapidly than $|\xi|^{-1}$. This should be contrasted with the computation of the Fourier transform of r_1. The function $\hat{r}_1$ tends to zero as $|\xi|$ tends to infinity exactly like $|\xi|^{-1}$. This is a reflection of the fact that r_1 is not everywhere differentiable, having jump discontinuities at ± 1.

If f has j integrable derivatives then, by repeatedly integrating by parts, we get additional formulæ for $\hat{f}$

$$\hat{f}(\xi) = \left[\frac{1}{i\xi}\right]^j \widehat{f^{[j]}}(\xi).$$

Again, because $f^{[j]}$ is absolutely integrable $\widehat{f^{[j]}}$ tends to zero as $|\xi| \to \infty$. We state the result of these computations as a proposition.

Proposition 4.2.1. *Let j be a positive integer. If f has j integrable derivatives then there is a constant C so $\hat{f}$ satisfies the estimate*

$$|\hat{f}(\xi)| \leq \frac{C}{(1+|\xi|)^j}.$$

Moreover, for $1 \leq l \leq j$, the Fourier transform of $f^{[l]}$ is given by

$$\widehat{f^{[l]}}(\xi) = (i\xi)^l \hat{f}(\xi). \tag{4.24}$$

The rate of decay in $\hat{f}$ is also reflected in the smoothness of f.

Proposition 4.2.2. *Suppose that f is absolutely integrable and j is a non-negative integer. If $\hat{f}$ decays like $|\xi|^{-(j+1+\epsilon)}$, for an $\epsilon > 0$, then f is continuous and has j continuous derivatives.*

Proof. The hypotheses of the proposition show that we may use the Fourier inversion formula to obtain

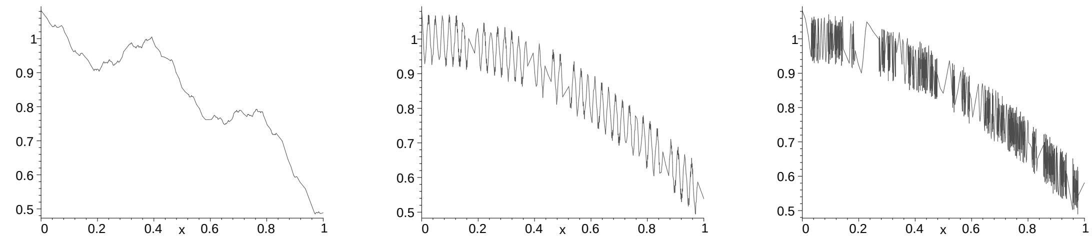
$$f(x) = \frac{1}{2\pi}\int\limits_{-\infty}^{\infty} \hat{f}(\xi)e^{ix\xi}d\xi.$$

In light of the decay of $\hat{f}$, we may differentiate this formula up to j times obtaining formulæ for derivatives of f as absolutely convergent integrals:

$$f^{[l]}(x) = \frac{1}{2\pi}\int\limits_{-\infty}^{\infty} \hat{f}(\xi)[i\xi]^l e^{ix\xi}d\xi \qquad \text{for } 0 \leq l \leq j.$$

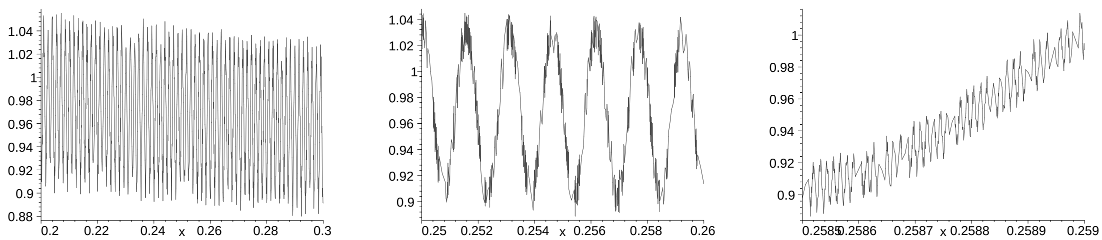
As $[i\xi]^l\hat{f}(\xi)$ is absolutely integrable for $l \leq j$ this shows that f has j continuous derivatives. ☐

Remark 4.2.3. Note that if f has j integrable derivatives then $\hat{f}$ decays faster than $|\xi|^{-j}$. The exact rate of decay depends on how continuous $f^{[j]}$ is. On the other hand, we need to assume that $\hat{f}$ decays *faster than* $|\xi|^{-(1+j)}$ to deduce that f has j-continuous derivatives. So we appear to "lose" one order of differentiability when inverting the Fourier transform. Both results are actually correct. The function r_1 provides an example showing the second result is sharp. It has a jump discontinuity and its Fourier transform, $2\operatorname{sinc}(\xi)$ decays like $|\xi|^{-1}$. To construct an example shoiwing that the first result is sharp, we now consider the case (not actually covered by the proposition) of $j = 0$. By integrating the examples constructed below, we obtain integrable functions with an integrable derivative whose Fourier transforms decay slower than $|\xi|^{-(1+\epsilon)}$, for any fixed positive $\epsilon > 0$.



(a) $k = 4$ (b) $k = 8$ (c) $k = 12$

Figure 4.1: Fuzzy functions.



(a) (b) (c)

Figure 4.2: The function $\operatorname{Re} f_{12}$ at smaller scales.

Example 4.2.5. Let φ be a smooth, rapidly decaying function with Fourier transform $\hat{\varphi}$ which satisfies the following conditions

(1). $0 \le \hat{\varphi}(\xi) \le 1$ for all ξ,

(2). $\hat{\varphi}(0) = 1$,

(3). $\hat{\varphi}(\xi) = 0$ if $|\xi| > 1$.

For example we could take

$$\hat{\varphi}(\xi) = \begin{cases} e^{-\frac{1}{1-\xi^2}} & \text{if } |\xi| < 1, \\ 0 & \text{if } |\xi| \ge 1. \end{cases}$$

In fact the details of this function are not important, only the listed properties are needed to construct the examples. For each $k \in \mathbb{N}$ define the function

$$\hat{f}_k(\xi) = \sum_{n=1}^{\infty} \frac{\hat{\varphi}(\xi - n^k)}{n^2}.$$

For a given ξ at most one term in the sum is non-zero. If $k > 1$ then $\hat{f}_k$ is zero "most of the time." On the other hand the best *rate of decay* that is holds for all ξ is

$$|\hat{f}_k(\xi)| \leq \frac{C}{|\xi|^{\frac{2}{k}}}.$$

By using a large k we can make this function decay as slowly as we like. Because $\sum n^{-2} < \infty$ the Fourier inversion formula applies to give

$$f_k(x) = \varphi(x) \sum_{n=1}^{\infty} \frac{e^{ixn^k}}{n^2}.$$

The infinite sum converges absolutely and uniformly in x and therefore f_k is a continuous function. Because φ decays rapidly at infinity so does f_k. This means that f_k is an absolutely integrable, continuous function whose Fourier transform goes to zero like $|\xi|^{-\frac{2}{k}}$. These examples show that the rate of decay of the Fourier transform of a continuous, absolutely integrable function can be as slow as one likes. The function, f_k is the smooth function φ, *modulated* by noise. The graphs in figure 4.1 show the real parts of these functions; they are very "fuzzy." The fact that these functions are not differentiable is visible in figure 4.2. These graphs show f_{12} at smaller and smaller scales. Observe that f_{12} does not appear smoother at small scales than at large scales.

These examples demonstrate that there are two different phenomena governing the rate of decay of the Fourier transform. The function r_1 is very smooth, except where it has a jump. This kind of very localized failure of smoothness produces a characteristic $|\xi|^{-1}$ rate of decay in the Fourier transform. In the L^1-sense the function r_1 is very close to being a continuous function. In fact by using linear interpolation we can find piecewise differentiable functions very close to r_1. These sort of functions frequently arise in medical imaging. Each function f_k is continuous, but very fuzzy. The larger k is, the higher the amplitude of the high frequency components producing the fuzz and the slower the rate of decay for $\hat{f}_k$. These functions are not close to differentiable functions in the L^1 sense. Such functions are typical of random processes used to model noise.

The forgoing results establish the connection between the regularity of f and the decay of its Fourier transform. If on the other hand, we know that f itself decays then this is reflected in increased regularity of its Fourier transform.

Proposition 4.2.3. *Suppose that j is a positive integer and*

$$\int_{-\infty}^{\infty} |f(x)|(1+|x|)^j dx < \infty,$$

then $\hat{f}(\xi)$ has j-continuous derivatives which tend to zero as $|\xi|$ tends to infinity. In fact for $0 \leq k \leq j$

$$\partial_\xi^k \hat{f}(\xi) = \int_{-\infty}^{\infty} (-ix)^k f(x) e^{-ix\xi} dx. \tag{4.25}$$

Of course (4.25) gives a formula for the Fourier transform of $x^k f(x)$ in terms of the Fourier transform of f :

$$\mathcal{F}(x^k f)(\xi) = i^k \partial_\xi^k \hat{f}(\xi). \tag{4.26}$$

A special case of this proposition arises if f vanishes outside a bounded interval. In this case $x^k f(x)$ is absolutely integrable for any positive integer k and therefore $\hat{f}$ is an infinitely differentiable function. The derivatives tend to zero as $|\xi|$ tends to infinity but the *rate* of decay may be the same for all the derivatives, for example

$$\hat{r}_1(\xi) = \frac{2\sin\xi}{\xi}.$$

Differentiating this function repeatedly gives a sum of terms one of which tends to zero exactly like $|\xi|^{-1}$. This further confirms our principle that the rate of decay of the Fourier transform is a reflection of the smoothness of the function.

Example 4.2.6. An important application of the Fourier transform is to study ordinary differential equations with constant coefficients. Suppose that $\{a_0, \dots, a_n\}$ are complex numbers, we would like to study the solutions of the differential equation

$$Af \overset{d}{=} \sum_{j=0}^{n} a_j \partial_x^j f = g.$$

Proceeding formally, take the Fourier transform of both sides, (4.24) gives relation

$$\left[\sum_{j=0}^{n} a_j (i\xi)^j\right] \hat{f}(\xi) = \hat{g}(\xi) \tag{4.27}$$

The polynomial,

$$P_A(\xi) = \sum_{j=0}^{n} a_j (i\xi)^j$$

is called the *characteristic polynomial* for the differential operator A. If a complex number ξ_0 is a root of this equation, i.e. $P_A(\xi_0) = 0$ then the exponential function $v_0 = \exp(i\xi_0 x)$ is a solution of the homogeneous equation $Av_0 = 0$.

If on the other hand, P_A has no real roots and g is absolutely integrable then we can divide in (4.27) to obtain

$$\hat{f}(\xi) = \frac{\hat{g}(\xi)}{P_A(\xi)}.$$

Using the Fourier inversion formula we obtain a particular solution to the equation $Af = g$,

$$f_p(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\hat{g}(\xi) e^{i\xi x} d\xi}{P_A(\xi)}. \tag{4.28}$$

The general solution is of the form $f_p + f_0$ where $Af_0 = 0$. If P_A has real roots then a more careful analysis is required, see [7].

Exercises

Exercise 4.2.9. * Let f be an absolutely integrable function. Show that $\hat{f}$ is a continuous function. Extra credit: Show that $\hat{f}$ is uniformly continuous on the whole real line.

Exercise 4.2.10. If fg is absolutely integrable show that sequences exist satisfying (4.22).

Exercise 4.2.11. Suppose that $f'g$ and fg' are absolutely integrable. Show that the limits

$$\lim_{x\to\infty} fg(x) \text{ and } \lim_{x\to-\infty} fg(x)$$

both exist. Does (4.21) hold even if fg is not assumed to be absolutely integrable?

Exercise 4.2.12. Prove that for any number j the j^{th}-derivative $\partial_\xi^j \hat{r}_1$ has a term which decays exactly like $|\xi|^{-1}$.

Exercise 4.2.13. Show that f_p, defined in example 4.28 and its first n derivatives tend to zero as $|x|$ tends to infinity.

Exercise 4.2.14. Show that the function, φ defined in example 4.2.5 is infinitely differentiable.

4.2.5 Fourier transform on $L^2(\mathbb{R})$

See: A.2.4, A.2.5, A.5.2, A.5.5.

In the foregoing discussion we considered absolutely integrable functions. The Fourier transform is then defined in terms of an absolutely convergent integral. As we observed, this does not imply that the Fourier transform is itself absolutely integrable. In fact, it is quite difficult to describe the range of $\mathcal{F}$ when the domain is $L^1(\mathbb{R})$. Using the L^1-norm, there are also discrepancies in the quantitative relationships between the smoothness of a function and the rate of decay of its Fourier transform. A more natural condition, when working with Fourier transform is *square-integrability.*

Definition 4.2.4. A complex valued function f, defined on $\mathbb{R}^n$ is square-integrable if

$$\|f\|_{L^2}^2 = \int\limits_{\mathbb{R}^n} |f(\mathbf{x})|^2 d\mathbf{x} < \infty.$$

The set of such functions, with norm defined by $\|\cdot\|_{L^2}$, is denoted $L^2(\mathbb{R}^n)$. With this norm $L^2(\mathbb{R}^n)$ is a complete, normed linear space.

The norm on $L^2(\mathbb{R}^n)$ is defined by an inner product,

$$\langle f, g\rangle_{L^2} = \int\limits_{\mathbb{R}^n} f(\mathbf{x})\overline{g(\mathbf{x})} d\mathbf{x}.$$

This inner product satisfies the usual Cauchy-Schwarz inequality.

Proposition 4.2.4. *If $f, g \in L^2(\mathbb{R}^n)$ then*

$$|\langle f, g\rangle_{L^2}| \leq \|f\|_{L^2}\|g\|_{L^2}. \tag{4.29}$$

Proof. The proof of the Cauchy-Schwarz inequality for $L^2(\mathbb{R}^n)$ is formally identical to the proof for $\mathbb{C}^n$ given in the proof of Proposition 2.2.2. The verification of this fact is left to the exercises. □

Recall a normed linear space is complete if every Cauchy sequence has a limit. The completeness of L^2 is quite important for what follows.

Example 4.2.7. The function $f(x) = (1 + |x|)^{-\frac{3}{4}}$ is not absolutely integrable, but it is square-integrable. On the other hand the function

$$g(x) = \frac{\chi_{[-1,1]}(x)}{\sqrt{|x|}}$$

is absolutely integrable but not square-integrable.

An L^2-function is always *locally absolutely integrable.* This means that for any finite interval, $[a, b]$ the following integral is finite

$$\int_a^b |f(x)|dx.$$

To prove this we use the Cauchy-Schwarz inequality with $g = 1$:

$$\int_a^b |f(x)|dx \leq \sqrt{|b-a|}\sqrt{\int_a^b |f(x)|^2 dx} \leq \sqrt{|b-a|}\|f\|_{L^2}.$$

The reason square-integrability is a natural condition is contained in the following theorem.

Theorem 4.2.2 (Parseval formula). *If f is absolutely integrable and also square-integrable, then $\hat{f}$ is square-integrable and*

$$\int_{-\infty}^{\infty} |f(x)|^2 dx = \int_{-\infty}^{\infty} |\hat{f}(\xi)|^2 \frac{d\xi}{2\pi}. \tag{4.30}$$

Though very typical of arguments in this subject, the proof of this result is rather abstract. It can safely be skipped as nothing in the sequel relies upon it.

Proof. To prove (4.30) we use the Fourier inversion formula, Propositions 4.2.1 and 4.2.3 and the following lemma.

Lemma 4.2.2. *Suppose that f and g are integrable functions which are $O(|x|^{-2})$ as $|x|$ tend to infinity. Then we have the identity*

$$\int_{-\infty}^{\infty} f(x)\hat{g}(x)dx = \int_{-\infty}^{\infty} \hat{f}(x)g(x)dx. \tag{4.31}$$

The proof of the lemma is left as an exercise.

Suppose for the moment that f is an infinitely differentiable function with bounded support. Proposition 4.2.1 shows that, for any positive k, $\hat{f}$ is $O(|\xi|^{-k})$ as $|\xi|$ tends to infinity while Proposition 4.2.3 shows that $\hat{f}$ is smooth and similar estimates hold for its derivatives. Let $g = [2\pi]^{-1}\overline{\hat{f}}$; the Fourier inversion formula implies that $\hat{g} = \bar{f}$. The identity (4.31) applies to this pair giving

$$\int_{-\infty}^{\infty} |f(x)|^2 dx = \frac{1}{2\pi} \int_{-\infty}^{\infty} |\hat{f}(\xi)|^2 d\xi.$$

Thus verifying the Parseval formula for the special case of smooth functions with bounded support.

In Chapter 5 we show that, for any function satisfying the hypotheses of the theorem, there is a sequence $< f_n >$ of smooth functions, with bounded support such that

$$\lim_{n\to\infty} \|f - f_n\|_{L^1} = 0 \text{ and } \lim_{n\to\infty} \|f - f_n\|_{L^2} = 0.$$

The proof given so far applies to the differences to show that

$$\int_{-\infty}^{\infty} |f_n(x) - f_m(x)|^2 dx = \frac{1}{2\pi} \int_{-\infty}^{\infty} |\hat{f}_n(\xi) - \hat{f}_m(\xi)|^2 d\xi. \tag{4.32}$$

The left hand side tends to zero as m, n tend to infinity. Therefore $< \hat{f}_n >$ is also an L^2-Cauchy sequence converging to a limit in $L^2(\mathbb{R})$. As $< f_n >$ converges to f in $L^1(\mathbb{R})$ the sequence $< \hat{f}_n >$ converges pointwise to $\hat{f}$. This implies that $< \hat{f}_n >$ tends to $\hat{f}$ in $L^2(\mathbb{R})$ as well. This completes the proof of the theorem as

$$\int_{-\infty} |f(x)|^2 dx = \lim_{n\to\infty} \int_{-\infty} |f_n(x)|^2 dx = \lim_{n\to\infty} \frac{1}{2\pi} \int_{-\infty} |\hat{f}_n(\xi)|^2 d\xi = \frac{1}{2\pi} \int_{-\infty} |\hat{f}(\xi)|^2 d\xi.$$

□

In many physical applications the square-integral of a function is interpreted as the total energy. Up to the factor of 2π, Parseval's formula says that the total energy in f is the same as that in $\hat{f}$. Often the variable $\frac{\xi}{2\pi}$ is thought of as a frequency. Following the quantum mechanical practice, higher frequencies correspond to higher energies. In this context $|\hat{f}(\xi)|^2$ is interpreted as the energy *density* of f at frequency $\frac{\xi}{2\pi}$. As we shall see, "noise" is essentially a high frequency phenomenon, and a noisy signal has a lot of energy at high frequencies.

The Parseval formula shows that the L^2-norm is intimately connected to the Fourier transform. When the L^2-norm is used in both the domain and range, Parseval's formula says that $\mathcal{F}$ is a continuous linear transformation. This result indicates that it should be possible to extend the Fourier transform to all functions in $L^2(\mathbb{R})$. This is indeed the case. Let $f \in L^2(\mathbb{R})$, for each $R > 0$ define

$$\hat{f}_R(\xi) = \int_{-R}^{R} f(x) e^{-ix\xi} dx. \tag{4.33}$$

From Parseval's formula it follows that, if $R_1 < R_2$ then

$$\|\hat{f}_{R_1} - \hat{f}_{R_2}\|_{L^2}^2 = 2\pi \int_{R_1 \le |x| \le R_2} |f(x)|^2 dx.$$

Because f is square-integrable the right hand side of this formula goes to zero as R_1 and R_2 tend to infinity. This says that, if we measure the distance in the L^2-norm, then the functions $< \hat{f}_R >$ are clustering closer and closer together as $R \to \infty$. Otherwise put, $< \hat{f}_R >$ is an L^2-Cauchy sequence. Because $L^2(\mathbb{R})$ is a complete, normed vector space, this implies that $< \hat{f}_R >$ converges to a limit as $R \to \infty$; this limit *defines* $\hat{f}$. The limit of a sequence in the L^2-norm is called a *limit in the mean*; it is denoted by the symbol LIM.

Definition 4.2.5. If f is a function in $L^2(\mathbb{R})$ then its Fourier transform is defined to be

$$\hat{f} = \underset{R\to\infty}{LIM} \hat{f}_R,$$

where $\hat{f}_R$ is defined in (4.33).

We summarize these observations in a proposition.

Proposition 4.2.5. *The Fourier transform extends to define a continuous map from $L^2(\mathbb{R})$ to itself. If $f \in L^2(\mathbb{R})$ then*

$$\int_{-\infty}^{\infty} |f(x)|^2 dx = \frac{1}{2\pi} \int_{-\infty}^{\infty} |\hat{f}(\xi)|^2 d\xi.$$

Proof. The continuity statement follows from the Parseval formula. That the Parseval formula holds for all $f \in L^2(\mathbb{R})$ is a consequence of the definition of $\hat{f}$ and the fact that

$$\int_{-R}^{R} |f(x)|^2 dx = \frac{1}{2\pi} \int_{-\infty}^{\infty} |\hat{f}_R(\xi)|^2 d\xi \text{ for } R > 0.$$

□

In other words the Fourier transform extends to define a continuous map from $L^2(\mathbb{R})$ to itself. There is however a price to pay as the Fourier transform of a function in $L^2(\mathbb{R})$ cannot be directly defined by a simple formula like (4.3). For a function like f in example 4.2.7 the integral defining $\hat{f}$ is not absolutely convergent.

Example 4.2.8. The function

$$f(x) = \frac{1}{\sqrt{1+x^2}}$$

is square integrable but not absolutely integrable. We use integration by parts to compute $\hat{f}_R(\xi)$:

$$\hat{f}_R(\xi) = \frac{2\,\mathrm{sinc}(R\xi)}{\sqrt{1+R^2}} - \int_{-R}^{R} \frac{xe^{-ix\xi}}{i\xi(1+x^2)^{\frac{3}{2}}}.$$

It is now a simple matter to obtain the pointwise limit as R tends to infinity:

$$\hat{f}(\xi) = \int_{-\infty}^{\infty} \frac{xe^{-ix\xi}}{i\xi(1+x^2)^{\frac{3}{2}}}. \tag{4.34}$$

In the exercise 4.2.16 you are asked to show that $\hat{f}_R$ converges to $\hat{f}$ in the L^2-norm.

A consequence of Parseval's formula is the identity.

$$\int_{-\infty}^{\infty} f(x)\overline{g(x)}dx = \int_{-\infty}^{\infty} \hat{f}(\xi)\overline{\hat{g}(\xi)}\frac{d\xi}{2\pi}. \tag{4.35}$$

This is proved by applying (4.30) to $f + tg$ and comparing the coefficients of powers t on the right- and left-hand sides. Up to the factor of 2π, the Fourier transform preserves the inner product. Recall that this is also a property of rotations of Euclidean space. Such transformations of complex vector spaces are called *unitary.* Another consequence of the Parseval formula is a uniqueness statement: a function in L^2 is determined by its Fourier transform.

Corollary 4.2.1. *If $f \in L^2(\mathbb{R})$ and $\hat{f} = 0$, then $f \equiv 0$.*

Remark 4.2.4. As noted in section 4.2.1 it would be more accurate to say that the set of points for which $f \neq 0$ has measure 0.

The Fourier transform of a square-integrable function is generally not absolutely integrable so the inversion formula, proved above, does not directly apply. The inverse is defined in much the same way as the Fourier transform itself.

Proposition 4.2.6 (Fourier inversion for $L^2(\mathbb{R})$). *For $f \in L^2(\mathbb{R})$ define*

$$F_R(x) = \frac{1}{2\pi}\int_{-R}^{R} \hat{f}(\xi)d\xi,$$

then $f = \underset{R\to\infty}{LIM} F_R$.

The proof is left to the exercises.

We conclude this section by summarizing the basic properties of the Fourier transform which hold for integrable or square-integrable functions. These properties are consequences of elementary properties of the integral.

1. Linearity:

   The Fourier transform is a linear operation:

$$\mathcal{F}(f+g) = \mathcal{F}(f) + \mathcal{F}(g), \quad \mathcal{F}(\alpha f) = \alpha\mathcal{F}(f),\ \alpha \in \mathbb{C}.$$

2. Scaling:

   The Fourier transform of $f(ax)$, the function f dilated by $a \in \mathbb{R}$, is given by

$$\begin{aligned}\int_{-\infty}^{\infty} f(ax)e^{-i\xi x}dx &= \int_{-\infty}^{\infty} f(y)e^{-\frac{i\xi y}{a}}\frac{dy}{a} \\ &= \frac{1}{a}\hat{f}\left(\frac{\xi}{a}\right).\end{aligned} \tag{4.36}$$

3. Translation:

   Let f_t be the function f shifted by t, i.e. $f_t(x) = f(x-t)$. The Fourier transform of f_t is given by

$$
\begin{aligned}
\widehat{f_t}(\xi) &= \int_{-\infty}^{\infty} f(x-t)e^{-i\xi x}dx \\
&= \int f(y)e^{-i\xi(y+t)}dy \\
&= e^{-i\xi t}\hat{f}(\xi).
\end{aligned}
\tag{4.37}
$$

4. Reality:

   If f is a real valued function then its Fourier transform satisfies $\hat{f}(\xi) = \overline{\hat{f}(-\xi)}$. This shows that the Fourier transform of a real valued function is completely determined by its values for positive (or negative) frequencies.

Recall the following definitions.

Definition 4.2.6. A function f defined on $\mathbb{R}^n$ is *even* if $f(\mathbf{x}) = f(-\mathbf{x})$. A function f defined on $\mathbb{R}^n$ is *odd* if $f(\mathbf{x}) = -f(-\mathbf{x})$.

5. Evenness:

   If f is even then $\hat{f}$ is real valued. If f is odd then $\hat{f}$ takes purely imaginary values. If f is even its Fourier transform is given by the formula

$$
\hat{f}(\xi) = 2\int_0^{\infty} f(x)\cos(\xi x)dx. \tag{4.38}
$$

Exercises

Exercise 4.2.15. Give a detailed proof of Proposition 2.2.2. Explain the following statement: “The Cauchy-Schwarz inequality is a statement about the 2-dimensional subspaces of a vector space.”

Exercise 4.2.16. Prove that $\hat{f}_R$, defined in (4.34) converges to $\hat{f}$ in the L^2-norm.

Exercise 4.2.17. Let $f(x) = \chi_{[1,\infty)}(x)x^{-1}$. Using the method of example 4.2.8 compute the Fourier transform of f. Verify the convergence, in the L^2-norm, of $\hat{f}_R$ to $\hat{f}$.

Exercise 4.2.18. * Prove Proposition 4.2.6. Hint: Use the Parseval formula to estimate the difference $\|f - F_R\|^2_{L^2}$.

Exercise 4.2.19. Let $f, g \in L^2(\mathbb{R})$ by considering the functions $f + tg$ where $t \in \mathbb{C}$ show that the Parseval formula implies (4.35).

Exercise 4.2.20. * Prove Lemma 4.2.2 and show that (4.31) holds for any pair of functions in $L^2(\mathbb{R})$.

Exercise 4.2.21. Verify the statement that if $g = \overline{\hat{f}}$ then $\hat{g} = \bar{f}$.

Exercise 4.2.22. Show that a function $f \in L^2(\mathbb{R})$ is zero if and only if $\langle f, g\rangle = 0$ for every $g \in L^2(\mathbb{R})$. Use this fact and the formula in exercise 4.2.20 to show that $\mathcal{F}(L^2(\mathbb{R})) = L^2(\mathbb{R})$. Hint: If this were false then there would exist a non-zero function $g \in L^2(\mathbb{R})$ such that $\langle g, \hat{f}\rangle = 0$ for every $f \in L^2(\mathbb{R})$.

Exercise 4.2.23. Verify properties (4) and (5).

Exercise 4.2.24. Find a formula like (4.38) for the Fourier transform of an odd function.

Exercise 4.2.25. * Suppose that $m(\xi)$ is a bounded function. Show that the map from $L^2(\mathbb{R})$ to itself defined by $\mathcal{A}_m(f) = \mathcal{F}^{-1}(m\hat{f})$ is continuous.

4.2.6 A general principle in functional analysis

In the previous section we extended the definition of the Fourier transform to $L^2(\mathbb{R})$ by using the Plancherel formula *and* the completeness of $L^2(\mathbb{R})$. This is an example of a general principle in functional analysis and explains, in part, why *completeness* is such an important property for a normed linear space. As we will encounter this situation again, we pause for a moment to ennuciate this principle explicitly. Recall the following definition.

Definition 4.2.7. Let $(V, \|\cdot\|)$ be a normed linear space. A subspace S of V is *dense* if for every $\mathbf{v} \in V$ there is a sequence $< \mathbf{v}_k > \subset S$ such that

$$\lim_{k\to\infty} \|\mathbf{v} - \mathbf{v}_k\| = 0.$$

The general principle is that a bounded linear map, defined on a dense subset extends to the whole space.

Theorem 4.2.3. *Let $(V_1, \|\cdot\|_1)$ and $(V_2, \|\cdot\|_2)$ be normed, linear spaces and assume that V_2 is complete. Suppose that S_1 is a dense subspace of V_1 and A is a linear map from S_1 to V_2. If there exists a constant M such that*

$$\|A\mathbf{v}\|_2 \le M\|\mathbf{v}\|_1, \tag{4.39}$$

for all $\mathbf{v}$ in S_1 then A extends to define a linear map from V_1 to V_2, satisfying the same estimate.

Proof. Let $\mathbf{v}$ be an arbitrary point in V_1 and let $< \mathbf{v}_k >$ be a sequence contained in S_1 converging to $\mathbf{v}$. Because A is linear and S_1 is a subspace, (4.39) implies that

$$\|A(\mathbf{v}_j - \mathbf{v}_k)\|_2 \le M\|\mathbf{v}_j - \mathbf{v}_k\|_1.$$

This estimate shows that $< A\mathbf{v}_k >$ is a Cauchy sequence in V_2. From the completeness of V_2 we conclude that this sequence has a limit $\mathbf{u}$. Provisionally define $A\mathbf{v} = \mathbf{u}$. To show that $A\mathbf{v}$ is well defined we need to show that if $< \mathbf{v}'_k > \subset S_1$ is another sequence converging to $\mathbf{v}$ then $< A\mathbf{v}'_k >$ also converges to $\mathbf{u}$. Since the two sequences have the same limit the difference $\|\mathbf{v}_k - \mathbf{v}'_k\|_1$ converges to zero. The estimate (4.39) implies that

$$\|A\mathbf{v}_k - A\mathbf{v}'_k\|_2 \le M\|\mathbf{v}_k - \mathbf{v}'_k\|_1,$$

showing that the limit is well defined. The fact that the extended map is linear is left as an exercise. □

Exercises

Exercise 4.2.26. Show that the extension of A defined in the proof of Theorem 4.2.3 is linear.

Exercise 4.2.27. Show that the only dense subspace of a finite-dimensional normed linear space is the whole space.

4.3 Functions with weak derivatives

See: A.5.6.

If f and g are differentiable functions which vanish outside a bounded interval then the integration by parts formula states that

$$\int_{-\infty}^{\infty} f'(x)\overline{g(x)}dx = -\int_{-\infty}^{\infty} f(x)\overline{g'(x)}dx. \tag{4.40}$$

This formula suggests a way to extend the notion of differentiability to some functions which do not have an ordinary derivative. Suppose that f is a locally integrable function and there exists another locally integrable function f_1 such that, for *every* $\mathcal{C}^1$-function g which vanishes outside a bounded interval, we have the identity

$$\int_{-\infty}^{\infty} f_1(x)\overline{g(x)}dx = -\int_{-\infty}^{\infty} f(x)\overline{g'(x)}dx.$$

From the point of view any measurement defined by $\mathcal{C}^1$-functions with bounded support, the function f_1 *looks like* the derivative of f. If this condition holds then we say that f has a *weak derivative* and write $f' = f_1$. In this context the function g is called a *test function*. It is clear from the definition that a function which is differentiable in the ordinary sense is weakly differentiable and the two definitions of derivative agree. When it is important to make a distinction the derivative defined as a limit of difference quotients is called a *classical derivative*.

An important fact about weak derivatives is that they satisfy the fundamental theorem of calculus. If f_1 is the weak derivative of f then for any $a < b$ we have that

$$f(b) - f(a) = \int_a^b f_1(x)dx. \tag{4.41}$$

The proof of this statement is a bit more involved than one might expect. The basic idea is to use a sequence of test functions in (4.40) which converge to $\chi_{[a,b]}$. We return to this is Chapter 5.

The definition of weak derivative can be applied recursively to define higher order weak derivatives. Suppose a locally integrable function f has a weak derivative f'. If f' also has a weak derivative then we say that f has two weak derivatives with $f^{[2]} \stackrel{d}{=} (f')'$. More generally if f has j weak derivatives, $\{f', \dots, f^{[j]}\}$ and $f^{[j]}$ has a weak derivative then we say that f has $j+1$ weak derivatives. The usual notations are also used for weak derivatives, i.e. f', $f^{[j+1]}$, $\partial_x^j f$, etc.

It is easy to see from examples that a weak derivative can exist even when f does not have a classical derivative.

Example 4.3.1. The function

$$f(x) = \begin{cases} 0 & \text{if } |x| > 1, \\ |x+1| & \text{if } -1 \leq x \leq 0, \\ |x-1| & \text{if } 0 \leq x \leq 1 \end{cases}$$

does not have a classical derivative at $x = -1, 0$ and 1. However the function

$$g(x) = \begin{cases} 0 & \text{if } |x| > 1, \\ 1 & \text{if } -1 \leq x \leq 0, \\ -1 & \text{if } 0 \leq x \leq 1 \end{cases}$$

is the weak derivative of f.

Weak differentiability is well adapted to the Fourier transform. Suppose that f is an absolutely integrable function with a weak derivative which is also absolutely integrable. The functions $fe^{-ix\xi}$ and $f'e^{-ix\xi}$ are in $L^1(\mathbb{R})$ and therefore formula (4.21) applies to show that

$$\int_{-\infty}^{\infty} f(x)e^{-ix\xi}dx = \frac{1}{i\xi}\int_{-\infty}^{\infty} f'(x)e^{-ix\xi}dx.$$

Thus the Fourier transform of the weak derivative is related to that of the original function precisely as in the classically differentiable case.

The notion of weak derivative extends the concept of differentiability to a larger class of functions. In fact this definition can be used to define derivatives of *generalized functions*. This topic is discussed in Appendix A.5.6; the reader is urged to look over this section.

Exercises

Exercise 4.3.1. In example 4.3.1 prove that g is the weak derivative of f.

Exercise 4.3.2. Show that if f has a weak derivative then it is continuous.

Exercise 4.3.3. Show that if f has two weak derivatives then it has one classical derivative.

Exercise 4.3.4. Show that the definition for higher order weak derivatives is consistent: If f has j weak derivatives and $f^{[j]}$ has k weak derivatives then f has $k+j$ weak derivatives and

$$f^{[k+j]} = (f^{[j]})^{[k]}.$$

Exercise 4.3.5. Show that Proposition 4.2.1 remains true if f is assumed to have j absolutely integrable, *weak* derivatives.

4.3.1 Functions with L^2-derivatives*

A especially useful condition is for the weak derivative to belong to L^2.

Definition 4.3.1. Let $f \in L^2(\mathbb{R})$ we say that f has an L^2-*derivative* if f has a weak derivative which also belongs to $L^2(\mathbb{R})$.

A function in $L^2(\mathbb{R})$ which is differentiable in the ordinary sense and whose classical derivative belongs to $L^2(\mathbb{R})$ is also differentiable in the L^2-sense. Its classical and L^2-derivatives are equal. Using (4.41) it is not difficult to show that a function with an L^2-derivative is continuous. Applying the Cauchy-Schwarz inequality to the right hand side of (4.41) gives the estimate

$$|f(b) - f(a)| \leq \sqrt{|b-a|}\|f_1\|_{L^2}.$$

In other words

$$\frac{|f(b) - f(a)|}{\sqrt{|b-a|}} \leq \|f_1\|_{L^2}.$$

A function for which this ratio is bounded is called a *Hölder*-$\frac{1}{2}$ function. Such a function is said to have a *half a classical derivative*.

If $f \in L^2(\mathbb{R})$ has an L^2-derivative then the Fourier transform of f and f' are related just as they would be if f had a classical derivative

$$\widehat{f'}(\xi) = i\xi\hat{f}(\xi).$$

Moreover the Parseval identity carries over to give

$$\int_{-\infty}^{\infty} |f'(x)|^2 dx = \frac{1}{2\pi}\int_{-\infty}^{\infty} |\xi|^2|\hat{f}(\xi)|^2 d\xi.$$

On the other hand if $\xi\hat{f}(\xi)$ is square-integrable then one can show that f has an L^2-derivative and its Fourier transform is $i\xi\hat{f}(\xi)$. This is what was meant by the statement that the relationship between the smoothness of a function and the decay of the Fourier transform is very close when these concepts are defined with respect to the L^2-norm.

The higher L^2-derivatives are defined exactly as in the classical case. If $f \in L^2(\mathbb{R})$ has an L^2-derivative, and $f' \in L^2$ also has an L^2-derivative, then we say that f has two L^2-derivatives. This can be repeated to define all higher derivatives. A simple condition for a function $f \in L^2(\mathbb{R})$ to have j L^2-derivatives, is that there are functions $\{f_1, \dots, f_j\} \subset L^2(\mathbb{R})$ so that for every j-times differentiable function φ, vanishing outside a bounded interval and $1 \leq l \leq j$ we have that

$$\langle f, \varphi^{[l]}\rangle_{L^2} = (-1)^l\langle f_l, \varphi\rangle_{L^2}.$$

The function f_l is then the l^{th} L^2-derivative of f. Standard notations are also used for the higher L^2-derivatives, e.g. $f^{[l]}$, $\partial_x^l f$, etc.
The basic result about L^2-derivatives is.

Theorem 4.3.1. *A function $f \in L^2(\mathbb{R})$ has j L^2-derivatives if and only if $\xi^j\hat{f}(\xi)$ is in $L^2(\mathbb{R})$. In this case*

$$\widehat{f^{[l]}}(\xi) = (i\xi)^l\hat{f}(\xi), \tag{4.42}$$

moreover

$$\int_{-\infty}^{\infty} |f^{[l]}(x)|^2 dx = \frac{1}{2\pi} \int_{-\infty}^{\infty} |\xi|^{2l} |\hat{f}(\xi)|^2 d\xi. \tag{4.43}$$

Exercises

Exercise 4.3.6. Suppose that $f \in L^2(\mathbb{R})$ has an $L^2(\mathbb{R})$-derivative f'. Show that if f vanishes for $|x| > R$ then so does f'.

Exercise 4.3.7. Prove that if $f \in L^2(\mathbb{R})$ has an L^2-derivative then $\widehat{f'}(\xi) = i\xi\hat{f}(\xi)$. Hint: Use (4.35).

Exercise 4.3.8. Show that if f has an L^2-derivative then $\hat{f}$ is absolutely integrable. Conclude that f is a continuous function.

Exercise 4.3.9. Use the result of exercise 4.3.8 to prove that (4.41) holds under the assumption that f and f' are in $L^2(\mathbb{R})$.

4.3.2 Fractional derivatives and L^2-derivatives*

See: A.5.6 .

In the previous section we extended the notion of differentiability to functions which do not have a classical derivative. In the study of the Radon transform it turns out to be useful to have other generalizations of differentiability. We begin with a generalization of the classical notion of differentiability.

The basic observation is the following: a function f has a derivative if the difference quotients

$$\frac{f(x+h) - f(x)}{h}$$

have a limit as $h \to 0$. In order for this limit to exist it is clearly necessary that the ratios

$$\frac{|f(x+h) - f(x)|}{|h|}$$

be uniformly bounded, for small h. Thus the basic estimate satisfied by a continuously differentiable function is that the ratios

$$\frac{|f(x) - f(y)|}{|x - y|}$$

are locally, uniformly bounded. The function $f(x) = |x|$ shows that these ratios can be bounded without the function being differentiable. However, from the point of view of measurements such a distinction is very hard to make.

Definition 4.3.2. Let $0 \leq \alpha < 1$, we say that a function f, defined in an interval $[a, b]$, has an α^{th}-*classical derivative* if there is a constant M so that

$$\frac{|f(x) - f(y)|}{|x - y|^{\alpha}} \leq M, \tag{4.44}$$

for all $x, y \in [a, b]$. Such a function is also said to be *α-Hölder continuous.*

The same idea can be applied to functions with L^2-derivatives. Recall that an L^2-function has an L^2-derivative if and only if $\xi \hat{f}(\xi) \in L^2(\mathbb{R})$. This is just the estimate

$$\int_{-\infty}^{\infty} |\xi|^2 |\hat{f}(\xi)|^2 d\xi < \infty.$$

By analogy to the classical case we make the following definition.

Definition 4.3.3. A function $f \in L^2(\mathbb{R})$ has an α^{th} *L^2-derivative* if

$$\int_{-\infty}^{\infty} |\xi|^{2\alpha} |\hat{f}(\xi)|^2 d\xi < \infty. \tag{4.45}$$

There is no canonical way to define the "α^{th}-L^2-derivative operator." The following definition is sometimes useful. For $\alpha \in (0, 1)$ define the α^{th}-L^2-derivative to be

$$D_{\alpha} f = \underset{R \to \infty}{LIM} \frac{1}{2\pi} \int_{-R}^{R} |\xi|^{\alpha} \hat{f}(\xi) e^{i\xi x} d\xi.$$

This operation is defined precisely for those functions satisfying (4.45). Note that this definition with $\alpha = 1$ does *not* give the expected answer.

The relationship between these two notions of fractional differentiability is somewhat complicated. As shown in the previous section: a function with one L^2-derivative is Hölder-$\frac{1}{2}$. On the other hand, the function $f(x) = \sqrt{x}$ is Holder-$\frac{1}{2}$. That $(\sqrt{x})^{-1}$ is not square integrable shows that having half a classical derivative does not imply that a function has one L^2-derivative.

Exercises

Exercise 4.3.10. Suppose that f satisfies the estimate in (4.44) with an $\alpha > 1$. Show that f is constant.

4.4 Some refined properties of the Fourier transform*

See: A.4.1, A.4.3.

In this section we consider some properties of the Fourier transform that are somewhat less elementary than those considered so far. The first question we consider concerns the *pointwise* convergence of the inverse Fourier transform.

4.4.1 Localization principle

Let f be a function in either $L^1(\mathbb{R})$ or $L^2(\mathbb{R})$; for each $R > 0$ define

$$f_R(x) = \mathcal{F}^{-1}(\chi_{[-R,R]}\hat{f})(x) = \frac{1}{2\pi}\int_{-R}^{R} \hat{f}(\xi)e^{ix\xi}d\xi.$$

The function f_R can be expressed directly in terms of f by the formula

$$f_R(x) = \int_{-\infty}^{\infty} f(y)\frac{\sin(R(x-y))}{\pi(x-y)}dy. \tag{4.46}$$

If f is in $L^1(\mathbb{R})$ then (4.46) follows by inserting the definition of $\hat{f}$ in the integral defining f_R and interchanging the order of the integrations. While if f is $L^2(\mathbb{R})$ then (4.46) follows from the formula in exercise 4.2.20.

If $\hat{f}$ is absolutely integrable then Theorem 4.2.1 shows that $f(x)$ is the limit, as $R \to \infty$ of $f_R(x)$. If f is well enough behaved *near to* x then this is always the case, whether or not $\hat{f}$ (or for that matter f) is absolutely integrable. This is Riemann's famous localization principle for the Fourier transform.

Theorem 4.4.1 (Localization principle). *Suppose that f belongs to either $L^1(\mathbb{R})$ or $L^2(\mathbb{R})$. If f vanishes in an open interval containing x_0 then*

$$\lim_{R\to\infty} f_R(x_0) = 0.$$

Proof. The proof of this result is not difficult. The same proof works f is in L^1 or L^2. From (4.46) we obtain

$$\begin{aligned} f_R(x_0) &= \int_{-\infty}^{\infty} f(y)\frac{\sin(R(x_0-y))}{\pi(x_0-y)}dy \\ &= \int_{-\infty}^{\infty} [e^{iR(x_0-y)} - e^{-iR(x_0-y)}]\frac{f(y)}{2\pi i(x_0-y)}dy. \end{aligned} \tag{4.47}$$

Because f vanishes in an interval containing x_0, it follows that $f(y)(x_0-y)^{-1}$ is an absolutely integrable function. The conclusion of the theorem is therefore a consequence of the Riemann-Lebesgue lemma. □

Remark 4.4.1. In light of the linearity of the Fourier transform this result holds for any function f which can be written as a sum $f = f_1 + f_2$ where $f_p \in L^p(\mathbb{R})$. The set of such functions is denoted $L^1(\mathbb{R}) + L^2(\mathbb{R})$. This set is clearly a vector space.

This result has a simple corollary which makes clearer why it is called the "localization principle." Suppose that f and g are functions in $L^1(\mathbb{R}) + L^2(\mathbb{R})$ such that

(1). $\lim_{R\to\infty} g_R(x_0) = g(x_0)$ and

(2). $f(x) = g(x)$ for x in an interval containing x_0.

The second condition implies that $f(x)-g(x)=0$ in an interval containing x_0 and therefore

$$\begin{aligned} f(x_0)=g(x_0) &= \lim_{R\to\infty} g_R(x_0) \\ &= \lim_{R\to\infty} g_R(x_0) + \lim_{R\to\infty}(f_R(x_0)-g_R(x_0)) \\ &= \lim_{R\to\infty} f_R(x_0). \end{aligned} \tag{4.48}$$

In the second line we use Theorem 4.4.1. The Fourier inversion process is very sensitive to the local behavior of f. It is important to note that this result is special to one dimension. The analogous result is *false* for the Fourier transform in $\mathbb{R}^n$ if $n \geq 2$. This phenomenon is carefully analyzed in [57], see also section 4.5.5.

Exercises

Exercise 4.4.1. Give a complete derivation for (4.46) with f either integrable or square-integrable.

Exercise 4.4.2. Suppose that f is a square integrable function which is continuously differentiable for $x \in (a,b)$. Show that for every x in this interval $\lim_{R\to\infty} f_R(x) = f(x)$.

4.4.2 The Heisenberg uncertainty principle

In this section we study relationships between the $\operatorname{supp} f$ and $\operatorname{supp} \hat{f}$. The simplest such result states that if a function has bounded support then its Fourier transform cannot.

Proposition 4.4.1. *Suppose* $\operatorname{supp} f$ *is contained the bounded interval* $(-R,R)$ *if* $\hat{f}$ *also has bounded support then* $f \equiv 0$.

Proof. The radius of convergence of the series $\sum_0^\infty (-ix\xi)^j/j!$ is infinity, and it converges to $e^{-ix\xi}$, uniformly on bounded intervals. Combining this with the fact that f has bounded support, we conclude that we may interchange the integration with the summation to obtain

$$\begin{aligned} \hat{f}(\xi) &= \int_{-\infty}^{\infty} f(x)e^{-i\xi x}dx \\ &= \int_{-R}^{R} \sum_{j=0}^{\infty} f(x)\frac{(-ix\xi)^j}{j!}dx \\ &= \sum_{j=0}^{\infty} \frac{1}{j!}(-i\xi)^j \int_{-R}^{R} f(x)x^j dx. \end{aligned} \tag{4.49}$$

Since

$$\left| \int_{-R}^{R} f(x)x^j dx \right| \leq R^j \int_{-R}^{R} |f(x)|dx,$$

the terms of the series representing $\hat{f}(\xi)$ are bounded by the terms of a series having an infinite radius of convergence; the j^{th} term is bounded by

$$\frac{(R|\xi|)^j}{j!} \int_{-R}^{R} |f(x)|dx.$$

Therefore the series expansion for $\hat{f}(\xi)$ also has an infinite radius of convergence. This argument can be repeated to obtain the Taylor expansion of $\hat{f}(\xi)$ about an arbitrary ξ_0 :

$$\begin{aligned}\hat{f}(\xi) &= \int_{-R}^{R} e^{-i(\xi-\xi_0)x} f(x)e^{i\xi_0 x}dx \\ &= \int_{-R}^{R}\sum_{j=0}^{\infty}\frac{[-i(\xi-\xi_0)x]^j}{j!}f(x)e^{i\xi_0 x}dx \\ &= \sum_{j=0}^{\infty}\int_{-R}^{R}\frac{[-i(\xi-\xi_0)x]^j}{j!}f(x)e^{i\xi_0 x}dx \\ &= \sum_{j=0}^{\infty}\frac{[-i(\xi-\xi_0)]^j}{j!}\int_{-R}^{R} f(x)x^j e^{i\xi_0 x}dx.\end{aligned} \tag{4.50}$$

If we let $a_j^{\xi_0} = \int f(x)x^j e^{i\xi_0 x}dx$ then

$$\hat{f}(\xi) = \sum_{0}^{\infty} a_j^{\xi_0}\frac{[-i(\xi-\xi_0)]^j}{j!}.$$

As above this expansion is valid for all ξ.

Suppose there exists ξ_0 such that $\partial_\xi^j \hat{f}(\xi_0) = 0$ for all $j = 0, 1, \ldots$ Then $\hat{f}(\xi) \equiv 0$ since all the coefficients, $a_j^{\xi_0} = \partial_\xi^j \hat{f}(\xi_0)$ equal zero. This proves the proposition. □

Remark 4.4.2. The proof actually shows that if f is supported on a finite interval and all the derivatives of $\hat{f}$ vanishes at a single point then $f \equiv 0$.

This result indicates that one cannot obtain both arbitrarily good resolution and denoising simultaneously. A famous quantitative version of this statement is the Heisenberg uncertainty principle which we now briefly discuss using physical terms coming from quantum mechanics and probability theory. The latter subject is discussed in Chapter 12. In this context an L^2-function f describes the state of a particle. The *probability* of finding the particle in the interval $[a, b]$ is defined to be $\int_a^b |f(x)|^2 dx$. We normalize so that the total probability is 1. By the Parseval formula,

$$\int_{-\infty}^{\infty} |f(x)|^2 dx = \int_{-\infty}^{\infty} |\hat{f}(\xi)|^2 \frac{d\xi}{2\pi} = 1.$$

The expected value of the position of a particle is given by

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

By translating in x we can normalize f to make $E(x)$ zero. In physics, the Fourier transform of f describes the momentum of a particle. The expected value of the momentum is

$$E(\xi) = \int \xi|\hat{f}(\xi)|^2 \frac{d\xi}{2\pi}.$$

By replacing f by $e^{i\xi_0 x}f$ for an appropriate choice of ξ_0 we can also make $E(\xi)=0$. With these normalizations, the variance of the position and the momentum, $(\Delta x)^2$ and $(\Delta\xi)^2$, are given by

$$\begin{aligned}(\Delta x)^2 &= \int_{-\infty}^{\infty} x^2|f(x)|^2 dx,\\ (\Delta\xi)^2 &= \int_{-\infty}^{\infty} \xi^2|\hat{f}(\xi)|^2\frac{d\xi}{2\pi}.\end{aligned}$$

The Parseval formula implies that

$$(\Delta\xi)^2 = \int_{-\infty}^{\infty} |\partial_x f(x)|^2 dx.$$

The basic result is

Theorem 4.4.2 (The Heisenberg uncertainty principle). *If f and $\partial_x f$ belong to $L^2(\mathbb{R})$ then*

$$\int_{-\infty}^{\infty} |x|^2|f(x)|^2 dx \int_{-\infty}^{\infty} |\xi|^2|\hat{f}(\xi)|^2\frac{d\xi}{2\pi} \geq \frac{1}{4}\left[\int_{-\infty}^{\infty} |f(x)|^2\right]^2 dx. \tag{4.51}$$

Because the product of the variances has a lower bound, this means that we cannot localize the position and the momentum of a particle, arbitrarily well *at the same time.* The proof of this theorem is a simple integration by parts followed by an application of the Cauchy-Schwarz inequality.

Proof. If f decays sufficiently rapidly, we can integration by parts to obtain that

$$\begin{aligned}\int_{-\infty}^{\infty} xff_x dx &= \frac{1}{2}(xf^2)\Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty}\frac{1}{2}f^2 dx\\ &= -\frac{1}{2}\int_{-\infty}^{\infty} f^2.\end{aligned} \tag{4.52}$$

The Cauchy-Schwarz inequality implies that

$$\left|\int_{-\infty}^{\infty} xff_x dx\right| \leq \left[\int_{-\infty}^{\infty} x^2|f|^2 dx\right]^{\frac{1}{2}}\left[\int_{-\infty}^{\infty}|f_x|^2 dx\right]^{\frac{1}{2}}$$

Using (4.52), the Parseval formula and this estimate we obtain

$$\frac{1}{2}\int_{-\infty}^{\infty}|f|^2 dx \leq \left[\int_{-\infty}^{\infty} x^2|f|^2 dx\right]^{\frac{1}{2}}\left[\frac{1}{2\pi}\int_{-\infty}^{\infty}\xi^2|\hat{f}|^2 dx\right]^{\frac{1}{2}}. \tag{4.53}$$

□

With the expected position and momentum normalized to be zero, the variance in the position and momentum are given by

$$\Delta x = \left(\int_{-\infty}^{\infty} x^2 f^2\right)^{1/2} \text{ and } \Delta\xi = \left(\int_{-\infty}^{\infty} f_x^2\right)^{1/2}.$$

The estimate (4.53) is equivalent to $\Delta x \cdot \Delta\xi \geq \frac{1}{2}$. If a, b are non-negative numbers then the arithmetic mean-geometric inequality states that

$$ab \leq \frac{a^2+b^2}{2}.$$

Combining this with the Heisenberg uncertainty principle shows that

$$1 \leq (\Delta x)^2 + (\Delta\xi)^2.$$

That is

$$\int_{-\infty}^{\infty} f^2 dx \leq \int_{-\infty}^{\infty} x^2 f^2 + f_x^2 dx. \tag{4.54}$$

The inequality (4.54) becomes an *equality* if we use the Gaussian function $f(x) = e^{-\frac{x^2}{2}}$. A reason why the Gaussian is often used to smooth measured data is that it provides the optimal resolution (in the L^2-norm) for a given amount of de-noising.

Exercises

Exercise 4.4.3. Show that both (4.51) and (4.54) are *equalities* if $f = e^{-\frac{x^2}{2}}$. Can you show that the only functions for which this is true are multiples of f?

4.4.3 The Paley-Wiener theorem

In imaging applications one usually works with functions of bounded support. The question naturally arises whether it is possible to recognize such a function from its Fourier transform. There are a variety of theorems which relate the support of a function to properties of its Fourier transform. They go collectively by the name of Paley-Wiener theorems.

Theorem 4.4.3 (Paley-Wiener Theorem I). *A square-integrable function f satisfies $f(x) = 0$ for $|x| > L$ if and only if its Fourier transform $\hat{f}$ extends to be an analytic function in the whole complex plane that satisfies*

$$\begin{gathered}\int_{-\infty}^{\infty} |\hat{f}(\xi + i\tau)|^2 d\xi \leq M e^{2L|\tau|} \text{ for all } \tau \text{ and} \\ |\hat{f}(\xi + i\tau)| \leq \frac{M e^{L|\tau|}}{\sqrt{|\tau|}}\end{gathered} \tag{4.55}$$

Proof. The proof of the forward implication is elementary. The Fourier transform of f is given by an integral over a finite interval,

$$\hat{f}(\xi) = \int_{-L}^{L} f(x)e^{-ix\xi}d\xi. \tag{4.56}$$

The expression clearly makes sense if ξ is replaced by $\xi + i\tau$, and differentiating under the integral shows that $\hat{f}(\xi + i\tau)$ is a analytic function. The first estimate follows from the Parseval formula as $\hat{f}(\xi+i\tau)$ is the Fourier transform of the L^2-function $f(x)e^{-\tau x}$. Using the Cauchy Schwartz inequality we obtain

$$\begin{aligned} |\hat{f}(\xi + i\tau)| &= \left| \int_{-L}^{L} f(x)e^{-ix\xi - x\tau} dx \right| \\ &\leq \frac{e^{L|\tau|}}{\sqrt{|\tau|}} \sqrt{\int_{-L}^{L} |f(x)|^2 dx}; \end{aligned} \tag{4.57}$$

from which the estimate is immediate.

The proof of the converse statement is a little more involved; it uses the Fourier inversion formula and a change of contour. We outline of this argument, the complete justification for the change of contour can be found in [40]. Let $x > L > 0$, the Fourier inversion formula states that

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{f}(\xi)e^{ix\xi}d\xi.$$

Since $\hat{f}(z)e^{ixz}$ is an analytic function, satisfying appropriate estimates, we can shift the integration to the line $\xi + i\tau$ for any $\tau > 0$,

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{f}(\xi + i\tau)e^{-x\tau}e^{ix\xi}d\xi.$$

In light of the first estimate in (4.55), we obtain the bound

$$|f(x)| \leq Me^{(L-x)\tau}.$$

Letting τ tend to infinity shows that $f(x) = 0$ for $x > L$. A similar argument using $\tau < 0$ shows that $f(x) = 0$ if $x < -L$. □

For later applications we state a variant of this result whose proof can be found in [40].

Theorem 4.4.4 (Paley-Wiener II). *A function $f \in L^2(\mathbb{R})$ has an analytic extension $F(x + iy)$ to the upper half plane ($y > 0$) satisfying*

$$\begin{aligned} &\int_{-\infty}^{\infty} |F(x + iy)|^2 dx \leq M, \\ &\lim_{y \downarrow 0} \int_{-\infty}^{\infty} |F(x + iy) - f(x)|^2 = 0 \end{aligned} \tag{4.58}$$

if and only if $\hat{f}(\xi) = 0$ for $\xi < 0$.

4.4.4 The Fourier transform of generalized functions

See: A.5.6.

Initially the Fourier transform is defined for absolutely integrable functions, by an explicit formula (4.3). It is then extended, in definition (4.2.5), to L^2-functions by using its *continuity* properties. The Parseval formula implies that the Fourier transform is a continuous map from $L^2(\mathbb{R})$ to itself, indeed it is an invertible, isometry. For an L^2-function, the Fourier transform is **not** defined by an integral, nonetheless the Fourier transform on $L^2(\mathbb{R})$ shares all the important properties of the Fourier transform defined earlier for absolutely integrable functions.

It is reasonable to seek the largest class of functions to which the Fourier transform can be extended. In turns out that the answer is *not* a class of functions, but rather the generalized functions (or tempered distributions) defined in section A.5.6. In the discussion which follows we assume a familiarity with this section! The definition of the Fourier transform on generalized functions closely follows the pattern of the definition of the derivative of a generalized function, with the result again a generalized function. To accomplish this extension we need to revisit the definition of a generalized function. In section A.5.6 we gave the following definition:

Let $\mathcal{C}_c^\infty(\mathbb{R})$ denote infinitely differentiable functions defined on $\mathbb{R}$ which vanish outside of bounded sets. These are called *test functions*.

Definition 4.4.1. A generalized function on $\mathbb{R}$ is a linear function, l defined on the set of test functions such that there is a constant C and an integer k so that, for every $f \in \mathcal{C}_c^\infty(\mathbb{R})$ we have the estimate

$$|l(f)| \leq C \sup_{x\in\mathbb{R}} \left[(1+|x|)^k \sum_{j=0}^{k} |\partial_x^j f(x)| \right] \tag{4.59}$$

These are linear functions on $\mathcal{C}_c^\infty(\mathbb{R})$ which are, in a certain sense continuous. The constants C and k in (4.59) depend on l but not on f. The expression on the right hand side defines a norm on $\mathcal{C}_c^\infty(\mathbb{R})$, for convenience we let

$$\|f\|_k = \sup_{x\in\mathbb{R}} \left[(1+|x|)^k \sum_{j=0}^{k} |\partial_x^j f(x)| \right].$$

The observation that we make is the following: if a generalized function satisfies the estimate

$$|l(f)| \leq C\|f\|_k$$

then it can be extended, by continuity, to any function f which is the limit of a sequence $< f_n > \subset \mathcal{C}_c^\infty(\mathbb{R})$ in the sense that

$$\lim_{n\to\infty} \|f - f_n\|_k = 0.$$

Clearly $f \in \mathcal{C}^k(\mathbb{R})$ and $\|f\|_k < \infty$. This motivates the following definition

Definition 4.4.2. A function $f \in \mathcal{C}^\infty(\mathbb{R})$ belongs to *Schwartz class* if $\|f\|_k < \infty$ for every $k \in \mathbb{N}$. The set of such functions is a vector space denoted by $\mathcal{S}(\mathbb{R})$.

From the definition it is clear that

$$\mathcal{C}_c^\infty(\mathbb{R}) \subset \mathcal{S}(\mathbb{R}). \tag{4.60}$$

Schwartz class does not have a norm with respect to which it is a complete normed linear space, instead each $\|\cdot\|_k$ defines a *semi-norm*. A sequence $< f_n > \subset \mathcal{S}(\mathbb{R})$ converges to $f \in \mathcal{S}(\mathbb{R})$ if and only if

$$\lim_{n\to\infty} \|f - f_n\|_k = 0 \text{ for every } k \in \mathbb{N}.$$

With this notion of convergence, Schwartz class becomes a complete metric space, the distance is defined by

$$d_{\mathcal{S}}(f,g) = \sum_{j=0}^{\infty} 2^{-j} \frac{\|f-g\|_j}{1+\|f-g\|_j}.$$

Remark 4.4.3. Of course each $\|\cdot\|_k$ satisfies all the axioms for a norm. They are called "semi-norms" because each one alone, does not define the topology on $\mathcal{S}(\mathbb{R})$.

Let $\varphi(x) \in \mathcal{C}_c^\infty(\mathbb{R})$ be a non-negative function with the following properties

(1). $\varphi(x) = 1$ if $x \in [-1,1]$,

(2). $\varphi(x) = 0$ if $|x| > 2$.

Define $\varphi_n(x) = \varphi(n^{-1}x)$, it is not difficult to prove the following proposition.

Proposition 4.4.2. *If $f \in \mathcal{S}(\mathbb{R})$ then $f_n = \varphi_n f \in \mathcal{C}_c^\infty(\mathbb{R}) \subset \mathcal{S}(\mathbb{R})$ converges to f in $\mathcal{S}(\mathbb{R})$. That is*

$$\lim_{n\to\infty} \|f_n - f\|_k = 0 \textit{ for every } k. \tag{4.61}$$

The proof is left as an exercise.

From the discussion above it therefore follows that *every* generalized function can be extended to $\mathcal{S}(\mathbb{R})$. Because (4.61) holds for every k, if l is a generalized function and $f \in \mathcal{S}(\mathbb{R})$ then $l(f)$ is defined as

$$l(f) = \lim_{n\to\infty} l(\varphi_n f).$$

To show that this makes sense, it is only necessary to prove that if $< g_n > \subset \mathcal{C}_c^\infty(\mathbb{R})$ which converges to f in Schwartz class then

$$\lim_{n\to\infty} l(g_n - \varphi_n f) = 0. \tag{4.62}$$

This is an immediate consequence of the triangle inequality and the estimate that l satisfies: there is a C and k so that

$$\begin{aligned} |l(g_n - \varphi_n f)| &\leq C\|g_n - \varphi_n f\|_k \\ &\leq C[\|g_n - f\|_k + \|f - \varphi_n f\|_k]. \end{aligned} \tag{4.63}$$

Since both terms on the right hand side of the second line tend to zero as $n \to \infty$, equation (4.62) is proved. In fact the generalized functions are exactly the set of continuous linear functions on $\mathcal{S}(\mathbb{R})$. For this reason the set of generalized functions is usually denoted by $\mathcal{S}'(\mathbb{R})$.

Why did we go to all this trouble? How will this help extend the Fourier transform to $\mathcal{S}'(\mathbb{R})$? The integration by parts formula was the "trick" used to extend the notion of derivative to generalized functions. The reason it works is that if $f \in \mathcal{S}(\mathbb{R})$ then $\partial_x f \in \mathcal{S}(\mathbb{R})$ as well. This implies that $l(\partial_x f)$ is a generalized function whenever l itself is. The Schwartz class has a similar property *vis á vis* the Fourier transform.

Theorem 4.4.5. *The Fourier transform is an isomorphism of $\mathcal{S}(\mathbb{R})$ onto itself, that is if $f \in \mathcal{S}(\mathbb{R})$ then both $\mathcal{F}(f)$ and $\mathcal{F}^{-1}(f)$ also belong to $\mathcal{S}(\mathbb{R})$. Moreover, for each k there is an k' and constant C_k so that*

$$\|\mathcal{F}(f)\|_k \leq C_k \|f\|_{k'} \text{ for all } f \in \mathcal{S}(\mathbb{R}). \tag{4.64}$$

The proof of this theorem is an easy consequence of results in section 4.2.4. We give the proof for $\mathcal{F}$, the proof for $\mathcal{F}^{-1}$ is essentially identical.

Proof. Since $f \in \mathcal{S}(\mathbb{R})$ for any $j, k \in \mathbb{N} \cup \{0\}$ we have the estimates

$$|\partial_x^j f(x)| \leq \frac{\|f\|_k}{(1+|x|)^k}. \tag{4.65}$$

From Propositions 4.2.1 and 4.2.3 it follows that $\hat{f}$ is infinitely differentiable and that, for any k, j,

$$\sup_{\xi \in \mathbb{R}} |\xi|^k |\partial_\xi^j \hat{f}(\xi)| < \infty.$$

To prove this we use the formula

$$\xi^k \partial_\xi^j \hat{f}(\xi) = \int_{-\infty}^{\infty} (i\partial_x)^k \left[(-ix)^j f(x)\right] e^{-ix\xi} dx.$$

Because $f \in \mathcal{S}(\mathbb{R})$ the integrand is absolutely integrable and in fact if $m = \max\{j, k\}$ then

$$|\xi^k \partial_\xi^j \hat{f}(\xi)| \leq C_{k,l} \|f\|_{m+2}, \tag{4.66}$$

here $C_{k,l}$ depends only on k and l. This completes the proof. □

Instead of integration by parts, we now use this theorem and the identity

$$\int_{-\infty}^{\infty} f(x)\hat{g}(x)dx = \int_{-\infty}^{\infty} \hat{f}(x)g(x)dx, \tag{4.67}$$

to extend the Fourier transform to generalized functions. The identity follows by a simple change in the order of integrations which is easily justified if $f, g \in \mathcal{S}(\mathbb{R})$. It is now clear how we should define the Fourier transform of a generalized function.

Definition 4.4.3. If $l \in \mathcal{S}'(\mathbb{R})$ then the Fourier transform of l is the generalized function $\hat{l}$ *defined* by

$$\hat{l}(f) = l(\hat{f}) \text{ for all } f \in \mathcal{S}(\mathbb{R}). \tag{4.68}$$

Theorem 4.4.5 implies that $\hat{f} \in \mathcal{S}(\mathbb{R})$ so that the right hand side in (4.68) defines a generalized function.

But why did we need to extend the definition of generalized functions from $\mathcal{C}_c^\infty(\mathbb{R})$ to $\mathcal{S}(\mathbb{R})$? The answer is simple: if $0 \neq f \in \mathcal{C}_c^\infty(\mathbb{R})$ then Proposition 4.4.1 implies that $\hat{f} \notin \mathcal{C}_c^\infty(\mathbb{R})$. This would prevent using (4.68) to define $\hat{l}$ because we would not know that $l(\hat{f})$ made sense! This appears to be a rather abstract definition and it is not at all clear that it can be used to compute the Fourier transform of a generalized function. In fact, there are many distributions whose Fourier transforms can be explicitly computed.

Example 4.4.1. If φ is an absolutely integrable function then

$$\widehat{l_\varphi} = l_{\hat{\varphi}}.$$

If $f \in \mathcal{S}(\mathbb{R})$ then the identity in (4.67) holds with $g = \hat{\varphi}$, as a simple interchange of integrations shows. Hence, for all $f \in \mathcal{S}(\mathbb{R})$

$$l_\varphi(\hat{f}) = \int_{-\infty}^{\infty} f(x)\hat{\varphi}(x)dx = l_{\hat{\varphi}}(f).$$

This shows that the Fourier transform for generalized functions is indeed an extension of the ordinary transform: if a generalized function l is *represented* by an integrable function in the sense that $l = l_\varphi$ then the definition of the Fourier transform of l is consistent with the earlier definition of the Fourier transform of φ.

Example 4.4.2. If $f \in \mathcal{S}(\mathbb{R})$ then

$$\hat{f}(0) = \int_{-\infty}^{\infty} f(x)dx.$$

This shows that $\hat{\delta} = l_1$ which is *represented* by an ordinary function equal to the constant 1.

Example 4.4.3. On the other hand the Fourier inversion formula implies that

$$\int_{-\infty}^{\infty} \hat{f}(\xi)d\xi = 2\pi f(0)$$

and therefore $\widehat{l_1} = 2\pi\delta$. This is an example of an ordinary function that does not have a Fourier transform, in the usual sense, and whose Fourier transform, as a generalized function is **not** an ordinary function.

Recall that a sequence $< l_n > \subset \mathcal{S}'(\mathbb{R})$ converges to l in $\mathcal{S}'(\mathbb{R})$ provided that

$$l(g) = \lim_{n\to\infty} l_n(g) \text{ for all } g \in \mathcal{S}(\mathbb{R}). \tag{4.69}$$

This is very useful for computing Fourier transforms because the Fourier transform is continuous with respect to the limit in (4.69). It follows from the definition that:

$$\widehat{l_n}(g) = l_n(\hat{g}) \tag{4.70}$$

and therefore

$$\lim_{n\to\infty} \widehat{l_n}(g) = \lim_{n\to\infty} l_n(\hat{g}) = l(\hat{g}) = \hat{l}(g). \tag{4.71}$$

Example 4.4.4. The generalized function $l_{\chi_{[0,\infty)}}$ can be defined as a limit by

$$l_{\chi_{[0,\infty)}}(f) = \lim_{\epsilon\downarrow 0} \int_0^\infty e^{-\epsilon x} f(x)dx.$$

The Fourier transform of $l_{e^{-\epsilon x}\chi_{[0,\infty)}}$ is easily computed using example 4.4.1, it is

$$\mathcal{F}(l_{e^{-\epsilon x}\chi_{[0,\infty)}})(f) = \int_{-\infty}^\infty \frac{f(x)dx}{ix+\epsilon}.$$

This shows that

$$\mathcal{F}(l_{\chi_{[0,\infty)}})(f) = \lim_{\epsilon\downarrow 0} \int_{-\infty}^\infty \frac{f(x)dx}{ix+\epsilon}. \tag{4.72}$$

In fact, it proves that the limit on the right hand side exists!

We close this discussion by verifying that the Fourier transform on generalized functions has many of the properties of the ordinary Fourier transform. Recall that if l is a generalized function and f is an infinitely differentiable function which satisfies estimates

$$|\partial_x^j f(x)| \leq C_j(1+|x|)^k,$$

for a *fixed* k then the product $f \cdot l$ is defined by

$$f \cdot l(g) = l(fg).$$

If $l \in \mathcal{S}'(\mathbb{R})$ then so are all of its derivatives. Using the definition it is not difficult to find formulæ for $\mathcal{F}(l^{[j]})$:

$$\mathcal{F}(l^{[j]})(f) = l^{[j]}(\hat{f}) = (-1)^j l(\partial_x^j \hat{f}) = l(\widehat{(ix)^j f}). \tag{4.73}$$

This shows that

$$\mathcal{F}(l^{[j]}) = (ix)^j \cdot \hat{l}. \tag{4.74}$$

A similar calculation shows that

$$\mathcal{F}((-ix)^j \cdot l) = \hat{l}^{[j]}. \tag{4.75}$$

Exercises

Exercise 4.4.4. Prove (4.60).

Exercise 4.4.5. Prove that $d_{\mathcal{S}}$ defines a metric. Show that a sequence $< f_n >$ converges in $\mathcal{S}(\mathbb{R})$ to f if and only if

$$\lim_{n\to\infty} d_{\mathcal{S}}(f_n, f) = 0.$$

Exercise 4.4.6. Prove Proposition 4.4.2.

Exercise 4.4.7. Prove (4.67). What is the "minimal" hypothesis on f and g so this formula makes sense, as absolutely convergent integrals.

Exercise 4.4.8. Give a detailed proof of (4.66).

Exercise 4.4.9. Prove, by direct computation that the limit on the right hand side of (4.72) exists for any $f \in \mathcal{S}(\mathbb{R})$.

Exercise 4.4.10. If $l_{1/x}$ is the Cauchy principal value integral

$$l_{1/x}(f) = \text{P. V.} \int_{-\infty}^{\infty} \frac{f(x)dx}{x}$$

then show that $\mathcal{F}(l_{1/x}) = l_{\operatorname{sign} x}$.

Exercise 4.4.11. Prove (4.75).

Exercise 4.4.12. The inverse Fourier transform of a generalized function is defined by

$$[\mathcal{F}^{-1}(l)](g) = l(\mathcal{F}^{-1}(g)).$$

Show that $\mathcal{F}^{-1}(\hat{l}) = l = \widehat{\mathcal{F}^{-1}(l)}$.

4.5 The Fourier transform for functions of several variables.

The Fourier transform can also be defined for functions of several variables. This section presents the definition and some of the elementary properties of the Fourier transform for functions in $L^1(\mathbb{R}^n)$ and $L^2(\mathbb{R}^n)$. In most ways it is quite similar to the one dimensional theory. A notable difference are discussed in section 4.5.5.

Recall that we use lower case, bold Roman letters $\mathbf{x}, \mathbf{y}$, etc. to denote points in $\mathbb{R}^n$, that is

$$\mathbf{x} = (x_1, \ldots, x_n) \text{ or } \mathbf{y} = (y_1, \ldots, y_n).$$

In this case x_j is called the j^{th}-coordinate of $\mathbf{x}$. The Fourier transform of a function of n-variables is also a function of n-variables. It is customary to use the lower case, bold Greek letters, $\boldsymbol{\xi}$ or $\boldsymbol{\eta}$ for points on the Fourier transform space with

$$\boldsymbol{\xi} = (\xi_1, \ldots, \xi_n) \text{ or } \boldsymbol{\eta} = (\eta_1, \ldots, \eta_n).$$

The volume form on Fourier space is denoted $d\boldsymbol{\xi} = d\xi_1 \ldots d\xi_n$.

4.5.1 L^1-case

As before we begin with the technically simpler case of absolutely integrable functions.

Definition 4.5.1. If f belongs to $L^1(\mathbb{R}^n)$ then the Fourier transform, $\hat{f}$ of f is defined by

$$\hat{f}(\boldsymbol{\xi}) = \int_{\mathbb{R}^n} f(\mathbf{x}) e^{-i\langle \boldsymbol{\xi}, \mathbf{x} \rangle} d\mathbf{x} \quad \text{for} \quad \boldsymbol{\xi} \in \mathbb{R}^n. \tag{4.76}$$

Since f is absolutely integrable over $\mathbb{R}^n$ the integral can be computed as an iterated integral

$$\int\limits_{\mathbb{R}^n} f(\mathbf{x})e^{-i\langle\boldsymbol{\xi},\mathbf{x}\rangle}d\mathbf{x} = \int\limits_{-\infty}^{\infty}\cdots\int\limits_{-\infty}^{\infty} f(x_1,\ldots,x_n)e^{-ix_1\xi_1}dx_1\cdots e^{-ix_n\xi_n}dx_n; \tag{4.77}$$

changing the order of the one dimensional integrals does not change the result. When thought of as a linear mapping, it is customary to use $\mathcal{F}(f)$ to denote the Fourier transform of f.

Using a geometric picture for the inner product leads to a better understanding of the functions $e^{i\langle\boldsymbol{\xi},\mathbf{x}\rangle}$. To that end we write $\boldsymbol{\xi}$ in polar form as $\boldsymbol{\xi} = r\boldsymbol{\omega}$. Here $r = \|\boldsymbol{\xi}\|$ is the length of $\boldsymbol{\xi}$ and $\boldsymbol{\omega}$ its direction. Write $\mathbf{x} = \mathbf{x}' + t\boldsymbol{\omega}$ where $\mathbf{x}'$ is orthogonal to $\boldsymbol{\omega}$, (i.e. $\langle\mathbf{x}',\boldsymbol{\omega}\rangle = 0$). As $\langle\mathbf{x},\boldsymbol{\omega}\rangle = t$ the function $\langle\mathbf{x},\boldsymbol{\omega}\rangle$ depends only t. Thus

$$e^{i\langle\mathbf{x},\boldsymbol{\xi}\rangle} = e^{irt}.$$

This function oscillates in the $\boldsymbol{\omega}$-direction with wave length $\frac{2\pi}{r}$. To illustrate this we give a density plot in the plane of the real and imaginary parts of

$$e^{i\langle\mathbf{x},\boldsymbol{\xi}\rangle} = \cos\langle\mathbf{x},\boldsymbol{\xi}\rangle + i\sin\langle\mathbf{x},\boldsymbol{\xi}\rangle$$

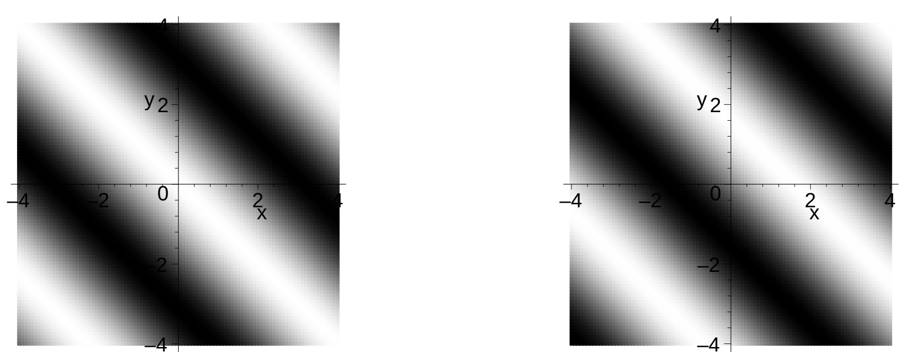
for several choices of $\boldsymbol{\xi}$. In these figures white corresponds to $+1$ and black corresponds to -1. The Fourier transform at $\boldsymbol{\xi} = r\boldsymbol{\omega}$ can be re-expressed as

$$\hat{f}(r\boldsymbol{\omega}) = \int\limits_{-\infty}^{\infty}\int\limits_{L} f(\mathbf{x}' + t\boldsymbol{\omega})e^{-irt}d\mathbf{x}'dt. \tag{4.78}$$

Here L is the $(n-1)$-dimensional subspace of $\mathbb{R}^n$ orthogonal to $\boldsymbol{\omega}$:

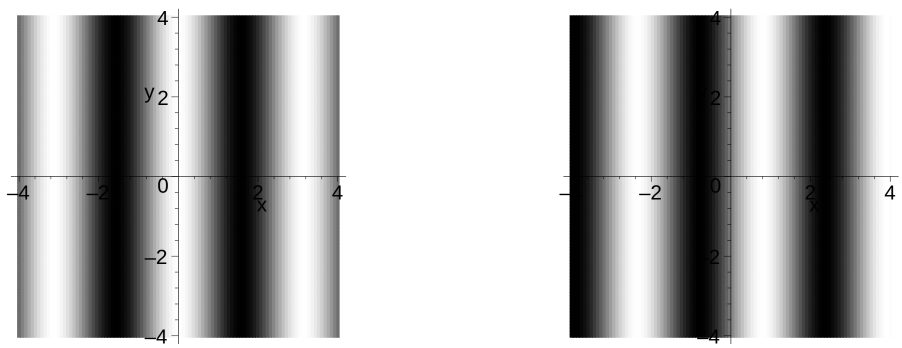
$$L = \{x' \in \mathbb{R}^n \ : \ \langle\mathbf{x}',\boldsymbol{\omega}\rangle = 0\}$$

and $d\mathbf{x}'$ is the $(n-1)$-dimensional Euclidean measure on L.



(a) (b)

Figure 4.3: Real and imaginary parts of $\exp(i\langle(x,y),(1,1)\rangle)$



(a) (b)

Figure 4.4: Real and imaginary parts of $\exp(i\langle(x,y),(2,0)\rangle)$

The Fourier transform is invertible; under appropriate hypotheses there is an explicit formula for the inverse.

Theorem 4.5.1 (Fourier Inversion Formula). *Suppose that f is an absolutely integrable function defined on $\mathbb{R}^n$. If $\hat{f}$ also belongs to $L^1(\mathbb{R}^n)$ then*

$$f(\mathbf{x}) = \frac{1}{[2\pi]^n}\int\limits_{\mathbb{R}^n} \hat{f}(\boldsymbol{\xi})e^{i\mathbf{x}\cdot\boldsymbol{\xi}}d\boldsymbol{\xi}. \tag{4.79}$$

Proof. The proof is formally identical to the proof of the one dimensional result. As before we begin by assuming that f is continuous. The basic fact used is that the Fourier transform of a Gaussian can be computed explicitly:

$$\mathcal{F}(e^{-\epsilon\|x\|^2}) = \left[\frac{\pi}{\epsilon}\right]^{\frac{n}{2}} e^{-\frac{\|x\|}{4\epsilon}}. \tag{4.80}$$

Because $\hat{f}$ is absolutely integrable

$$\begin{aligned}\frac{1}{[2\pi]^n}\int\limits_{\mathbb{R}^n} \hat{f}(\boldsymbol{\xi})e^{i\mathbf{x}\cdot\boldsymbol{\xi}}d\boldsymbol{\xi} &= \lim_{\epsilon\downarrow 0}\frac{1}{[2\pi]^n}\int\limits_{\mathbb{R}^n} \hat{f}(\boldsymbol{\xi})e^{i\mathbf{x}\cdot\boldsymbol{\xi}}e^{-\epsilon\|\boldsymbol{\xi}\|^2}d\boldsymbol{\xi}\\ &= \lim_{\epsilon\downarrow 0}\frac{1}{[2\pi]^n}\int\limits_{\mathbb{R}^n}\int\limits_{\mathbb{R}^n} f(\mathbf{y})e^{-i\mathbf{y}\cdot\boldsymbol{\xi}}d\mathbf{y}e^{i\mathbf{x}\cdot\boldsymbol{\xi}}e^{-\epsilon\|\boldsymbol{\xi}\|^2}d\boldsymbol{\xi}.\end{aligned} \tag{4.81}$$

The order of the integrations in the last line can be interchanged; using (4.80) gives,

$$\begin{aligned}\frac{1}{[2\pi]^n}\int\limits_{\mathbb{R}^n} \hat{f}(\boldsymbol{\xi})e^{i\mathbf{x}\cdot\boldsymbol{\xi}}d\boldsymbol{\xi} &= \lim_{\epsilon\downarrow 0}\frac{1}{[2\pi]^n}\int\limits_{\mathbb{R}^n} f(\mathbf{y})\left[\frac{\pi}{\epsilon}\right]^{\frac{n}{2}} e^{-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{4\epsilon}}d\mathbf{y}\\ &= \lim_{\epsilon\downarrow 0}\frac{1}{[2\pi]^{\frac{n}{2}}}\int\limits_{\mathbb{R}^n} f(\mathbf{x}-2\sqrt{\epsilon}\mathbf{t})e^{-\|\mathbf{t}\|^2}d\mathbf{t}.\end{aligned} \tag{4.82}$$

In the last line we use the change of variables $\mathbf{y} = \mathbf{x} - 2\sqrt{\epsilon}\mathbf{t}$. As f is continuous and absolutely integrable this converges to

$$\frac{f(\mathbf{x})}{[2\pi]^{\frac{n}{2}}}\int\limits_{\mathbb{R}^n} e^{-\|\mathbf{t}\|^2}d\mathbf{t}.$$

Since

$$\int_{\mathbb{R}^n} e^{-\|\mathbf{t}\|^2} d\mathbf{t} = [2\pi]^{\frac{n}{2}},$$

this completes the proof of the theorem for continuous functions. As in the one-dimensional case, an approximation argument is used to remove the additional hypothesis. The details are left to the reader. □

Exercises

Exercise 4.5.1. Prove formula (4.78).

Exercise 4.5.2. If $g_1(x), \dots, g_n(x)$ belong to $L^1(\mathbb{R})$ show that

$$f(x_1, \dots, x_n) = g_1(x_1) \cdots g_n(x_n) \in L^1(\mathbb{R}^n).$$

Show that

$$\hat{f}(\xi_1, \dots, \xi_n) = \hat{g}_1(\xi_1) \cdots \hat{g}_n(\xi_n).$$

Use this to compute the Fourier transform of $e^{-\|\mathbf{x}\|^2}$.

4.5.2 Regularity and decay

There is once again a close connection between the smoothness of function and the decay of its Fourier transform and vice versa. A convenient way to quantify the smoothness of a function on $\mathbb{R}^n$ is in terms of the existence of partial derivatives. Formulæ in several variables which involve derivatives can rapidly become cumbersome and unreadable. Fortunately there is a compact notation, called *multi-index* notation, giving n-variable formulæ with the simplicity and readability of the one-variable case.

Definition 4.5.2. A *multi-index* is an ordered n-tuple of non-negative integers usually denoted by a lower case Greek letter. For $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_n)$, a multi-index, set

$$\boldsymbol{\alpha}! = \alpha_1! \cdots \alpha_n! \qquad \text{and} \qquad |\boldsymbol{\alpha}| = \alpha_1 + \cdots + \alpha_n.$$

The function $|\boldsymbol{\alpha}|$ is called the length of $\boldsymbol{\alpha}$. The following conventions are useful:

$$\mathbf{x}^{\boldsymbol{\alpha}} = x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n} \text{ and } \partial_{\mathbf{x}}^{\boldsymbol{\alpha}} = \partial_{x_1}^{\alpha_1} \partial_{x_2}^{\alpha_2} \cdots \partial_{x_n}^{\alpha_n}.$$

Example 4.5.1. The binomial formula has an n-dimensional analogue:

$$(x_1 + \cdots + x_n)^k = k! \sum_{\{\boldsymbol{\alpha} : |\boldsymbol{\alpha}| = k\}} \frac{\mathbf{x}^{\boldsymbol{\alpha}}}{\boldsymbol{\alpha}!}.$$

Example 4.5.2. If f is a k-times differentiable function on $\mathbb{R}^n$ then there is also a n-dimensional analogue of Taylor's formula:

$$f(\mathbf{x}) = \sum_{\{\boldsymbol{\alpha} : |\boldsymbol{\alpha}| \leq k\}} \frac{\partial_{\mathbf{x}}^{\boldsymbol{\alpha}} f(0) \mathbf{x}^{\boldsymbol{\alpha}}}{\boldsymbol{\alpha}!} + R_k(\mathbf{x}). \tag{4.83}$$

Here R_k is the remainder term; it satisfies

$$\lim_{\|\mathbf{x}\| \to 0} \frac{|R_k(\mathbf{x})|}{\|\mathbf{x}\|^k} = 0.$$

As in the one dimensional case, the most general decay result is the Riemann-Lebesgue Lemma.

Proposition 4.5.1 (Riemann-Lebesgue Lemma). *Let f be an absolutely integrable function on $\mathbb{R}^n$ then $\hat{f}$ is a continuous function and* $\lim_{|\boldsymbol{\xi}|\to\infty}\hat{f}(\boldsymbol{\xi}) = 0$.

The proof is very similar to the one dimensional case and is left to the reader.

The smoothness of f is reflected in the decay properties of its Fourier transform. Suppose that f is continuous and has a continuous partial derivative in the x_j-direction which is integrable, i.e.

$$\int_{\mathbb{R}^n} |\partial_{x_j} f| d\mathbf{x} < \infty.$$

For notational convenience we suppose that $j = 1$ and set $\mathbf{x} = (x_1, \mathbf{x}')$. For any finite limits R, R_1, R_2 we can integrate by parts in the x_1-variable to obtain

$$\int_{-R_1}^{R_2} \int_{\|\mathbf{x}'\|<R} f(x_1, \mathbf{x}') e^{-i\langle \mathbf{x}, \boldsymbol{\xi}\rangle} d\mathbf{x}' dx_1 = \left[\frac{1}{-i\xi_1}\right] \left[\int_{\|\mathbf{x}'\|<R} [f(R_2, \mathbf{x}') - f(-R_1, \mathbf{x}')] e^{-i\langle \mathbf{x}, \boldsymbol{\xi}\rangle} d\mathbf{x}' - \int_{-R_1}^{R_2} \int_{\|\mathbf{x}'\|<R} \partial_{x_1} f(x_1, \mathbf{x}') e^{-i\langle \mathbf{x}, \boldsymbol{\xi}\rangle} d\mathbf{x}' dx_1 \right]. \tag{4.84}$$

Because $\partial_{x_1} f$ is integrable the second integral on the right hand side of (4.84) tends to $\mathcal{F}(\partial_{x_1} f)$ as R, R_1, R_2 tend to ∞. The boundary terms are bounded by

$$\int_{\mathbb{R}^{n-1}} [|f(R_2, \mathbf{x}')| + |f(-R_1, \mathbf{x}')|] d\mathbf{x}'.$$

As f is absolutely integrable there exist sequences $< R_1^k >, < R_2^k >$ tending to infinity so that these $(n-1)$-dimensional integrals tend to zero. This shows that $\mathcal{F}(\partial_{x_1} f) = i\xi_1 \mathcal{F}(f)$. The same argument applies to any coordinate proving the following proposition.

Proposition 4.5.2. *If f is an absolutely integrable, continuous function with an absolutely integrable, continuous j^{th}-partial derivative then*

$$\mathcal{F}(\partial_{x_j} f)(\boldsymbol{\xi}) = i\xi_j \mathcal{F}(f)(\boldsymbol{\xi}).$$

There is a constant C such that if f has a continuous, integrable gradient then $\hat{f}$ satisfies the estimate

$$|\hat{f}(\boldsymbol{\xi})| \leq \frac{C \int_{\mathbb{R}^n} [|f| + |\nabla f|] d\mathbf{x}}{(1 + \|\boldsymbol{\xi}\|)}.$$

The integration by parts argument can be iterated to obtain formulæ for the Fourier transform of $\partial_{\mathbf{x}}^{\boldsymbol{\alpha}} f$ for any multi-index $\boldsymbol{\alpha}$.

Theorem 4.5.2. *Suppose that f is an absolutely integrable, continuous function with continuous, absolutely integrable $\boldsymbol{\alpha}^{\text{th}}$-partial derivatives for any $\boldsymbol{\alpha}$ with $|\boldsymbol{\alpha}| < k$. Then there exists a constant C so that*

$$|\hat{f}(\boldsymbol{\xi})| \leq \frac{C}{(1+\|\boldsymbol{\xi}\|)^k},$$

and for each such $\boldsymbol{\alpha}$ we have

$$\widehat{\partial_{\mathbf{x}}^{\boldsymbol{\alpha}} f}(\boldsymbol{\xi}) = (i\boldsymbol{\xi})^{\boldsymbol{\alpha}} \hat{f}(\boldsymbol{\xi}).$$

The theorem relates the rate of decay of the Fourier transform to the smoothness of f. As in the one dimensional case, this theorem has a partial converse.

Proposition 4.5.3. *Suppose that f is an integrable function on $\mathbb{R}^n$ such that, for an $\epsilon > 0$, a non-negative integer k, and a constant C*

$$|\hat{f}(\boldsymbol{\xi})| \leq \frac{C}{(1+\|\boldsymbol{\xi}\|)^{k+n+\epsilon}} < \infty.$$

Then f has k-continuous derivatives which tend to zero at infinity.

Proof. The proof is a consequence of the Fourier inversion formula. The decay hypothesis implies that

$$f(\mathbf{x}) = \frac{1}{[2\pi]^n} \int_{\mathbb{R}^n} \hat{f}(\boldsymbol{\xi}) e^{i\mathbf{x}\cdot\boldsymbol{\xi}} d\boldsymbol{\xi}.$$

The estimate satisfied by $\hat{f}$ implies that this expression can be differentiated up to k-times. Hence the Fourier transform of $\partial_{\mathbf{x}}^{\boldsymbol{\alpha}} f$ is $(i\boldsymbol{\xi})^{\alpha} \hat{f}(\boldsymbol{\xi})$. Because $\mathcal{F}(\partial_{\mathbf{x}}^{\boldsymbol{\alpha}} f)$ is an L^1-function, the last statement follows from the Riemann-Lebesgue lemma. □

Remark 4.5.1. It is apparent that the discrepancy between this estimate and that in Theorem 4.5.2 grows as the dimension increases. As in the one-dimensional case, more natural and precise results are obtained by using weak derivatives and the L^2-norm. As these results are not needed in the rest of the book, we will not pursue this direction. The interested reader should consult [17].

In order to understand how decay at infinity for f is reflected in properties of $\hat{f}$ we first suppose that f vanishes outside the ball of radius R. It can be shown without difficulty that $\hat{f}(\boldsymbol{\xi})$ is a differentiable function, and its partial derivatives are given by

$$\partial_{\xi_j} \hat{f}(\boldsymbol{\xi}) = \int_{B_R} \partial_{\xi_j} [f(\mathbf{x}) e^{-i\boldsymbol{\xi}\cdot\mathbf{x}}] d\mathbf{x} = \int_{B_R} f(\mathbf{x})(-ix_j) e^{-i\boldsymbol{\xi}\cdot\mathbf{x}} d\mathbf{x} = \mathcal{F}(-ix_j f)(\boldsymbol{\xi}). \tag{4.85}$$

Iterating (4.85) gives

$$\partial_{\boldsymbol{\xi}}^{\boldsymbol{\alpha}} \hat{f}(\boldsymbol{\xi}) = (-i)^{|\boldsymbol{\alpha}|} \int_{\mathbb{R}^n} \mathbf{x}^{\boldsymbol{\alpha}} f(\mathbf{x}) e^{-i\boldsymbol{\xi}\cdot\mathbf{x}} d\mathbf{x} = (-i)^{|\boldsymbol{\alpha}|} \mathcal{F}(\mathbf{x}^{\boldsymbol{\alpha}} f)(\boldsymbol{\xi}). \tag{4.86}$$

If instead of assuming that f has bounded support, we assume that $(1+\|\mathbf{x}\|)^k f$ is integrable then a standard limiting argument shows that $\hat{f}$ is k times differentiable and the $\boldsymbol{\alpha}^{\text{th}}$ derivative is given by the right hand side of (4.86).

Summarizing these computations we have:

Proposition 4.5.4. *If $(1+\|\mathbf{x}\|)^k f$ is absolutely integrable for a positive integer k then the Fourier transform of f has k continuous derivatives. The partial derivatives of $\hat{f}$ are given by*

$$\partial_{\boldsymbol{\xi}}^{\boldsymbol{\alpha}} \hat{f}(\boldsymbol{\xi}) = (-i)^{|\boldsymbol{\alpha}|}\mathcal{F}(\mathbf{x}^{\boldsymbol{\alpha}} f)(\xi).$$

They satisfy the estimates

$$|\partial_{\boldsymbol{\xi}}^{\boldsymbol{\alpha}} \hat{f}(\boldsymbol{\xi})| \leq \int\limits_{\mathbb{R}^n} \|\mathbf{x}\|^{|\alpha|}|f(\mathbf{x})|d\mathbf{x},$$

and tend to zero as $\|\boldsymbol{\xi}\|$ tends to infinity.

Exercises

Exercise 4.5.3. Suppose that f is in $L^1(\mathbb{R}^n)$. Show that there exist sequences $< a_n >$ tending to $\pm\infty$ so that

$$\lim_{n\to\infty} \int\limits_{\mathbb{R}^{n-1}} |f(a_n, \mathbf{x}')|d\mathbf{x}'.$$

Exercise 4.5.4. Suppose that f is an integrable function which vanishes outside the ball of radius R. Show that $\hat{f}$ is a differentiable function and justify the interchange of the derivative and the integral in (4.85).

Exercise 4.5.5. Suppose that f is an integrable function which vanishes outside the ball of radius R. Show that $\hat{f}$ is an infinitely differentiable function.

Exercise 4.5.6. Give the details of the limiting argument used to pass from (4.86) with f of bounded support to the conclusion of Proposition 4.5.4.

Exercise 4.5.7. Prove the n-variable binomial formula.

Exercise 4.5.8. Explain the dependence on the dimension in the hypothesis of Proposition 4.5.3.

Exercise 4.5.9. Find a function f of n-variables so that

$$|\hat{f}(\boldsymbol{\xi})| \leq \frac{C}{(1+\|\boldsymbol{\xi}\|)^n}$$

but f is **not** continuous.

4.5.3 L^2-theory

See: A.5.6, A.5.5.

As in the 1-dimensional case, the n-dimensional Fourier transform extends to square-integrable functions. The basic result is the Parseval formula.

Theorem 4.5.3 (Parseval formula). *If f is absolutely integrable and $\int_{\mathbb{R}^n} |f(\mathbf{x})|^2 d\mathbf{x} < \infty$, then*

$$\int_{\mathbb{R}^n} |f(\mathbf{x})|^2 dx = \frac{1}{[2\pi]^n} \int_{\mathbb{R}^n} |\hat{f}(\boldsymbol{\xi})|^2 d\boldsymbol{\xi}. \tag{4.87}$$

The proof is quite similar to the one dimensional case. It uses an approximation argument and the identity, valid for absolutely integrable functions:

$$\int_{\mathbb{R}^n} f(\mathbf{x})\hat{g}(\mathbf{x}) d\mathbf{x} = \int_{\mathbb{R}^n} f(\mathbf{x}) \int_{\mathbb{R}^n} e^{-i\mathbf{x}\cdot\mathbf{y}} g(\mathbf{y}) d\mathbf{y} d\mathbf{x} = \int_{\mathbb{R}^n} \hat{f}(\mathbf{y}) g(\mathbf{y}) d\mathbf{y}. \tag{4.88}$$

The details are left to reader.

As in the single variable case the Fourier transform is extended to $L^2(\mathbb{R}^n)$ by continuity. As before we set

$$\hat{f}_R(\boldsymbol{\xi}) = \int_{\|\mathbf{x}\|<R} f(\mathbf{x}) e^{-i\boldsymbol{\xi}\cdot\mathbf{x}} d\mathbf{x}.$$

Parseval's formula implies that

$$\|\hat{f}_R\|_{L^2} \leq \|\chi_{B_R} f\|_{L^2}.$$

Because $L^2(\mathbb{R}^n)$ is complete and L^2-functions with bounded support are dense in L^2 it follows from Theorem 4.2.3 that the Fourier transform of f can be defined as the L^2-limit

$$\hat{f} = \underset{R\to\infty}{LIM} \hat{f}_R.$$

Moreover the Parseval formula extends to all functions in $L^2(\mathbb{R}^n)$. This shows that the Fourier transform is a continuous mapping of $L^2(\mathbb{R}^n)$ to itself: if $< f_n >$ is a sequence with $\underset{n\to\infty}{LIM} f_n = f$ then

$$\underset{n\to\infty}{LIM} \hat{f}_n = \hat{f}.$$

The L^2-inversion formula is also a consequence of the Parseval formula.

Proposition 4.5.5 (L^2-inversion formula). *Let $f \in L^2(\mathbb{R}^n)$ and define*

$$f_R(\mathbf{x}) = \frac{1}{[2\pi]^n} \int_{\|\boldsymbol{\xi}\|<R} \hat{f}(\boldsymbol{\xi}) e^{i\mathbf{x}\cdot\boldsymbol{\xi}} d\boldsymbol{\xi},$$

then $f = \underset{R\to\infty}{LIM} F_R$.

Proof. We need to show that $\lim_{R\to\infty} \|f_R - f\|_{L^2} = 0$. Because the norm is defined by an inner product we have

$$\|f_R - f\|_{L^2}^2 = \|f_R\|_{L^2}^2 - 2\operatorname{Re}\langle f_R, f\rangle_{L^2} + \|f\|_{L^2}.$$

The Parseval formula implies that

$$\|f_R\|_{L^2}^2 = \frac{1}{[2\pi]^n} \int_{\|\boldsymbol{\xi}\|<R} |\hat{f}(\boldsymbol{\xi})|^2 d\boldsymbol{\xi} \text{ and } \|f\|_{L^2}^2 = \frac{1}{[2\pi]^n} \int_{\mathbb{R}^n} |\hat{f}(\boldsymbol{\xi})|^2 d\boldsymbol{\xi}.$$

The proof is completed by using the following lemma.

Lemma 4.5.1. *Let $g \in L^2(\mathbb{R}^n)$ then*

$$\langle f_R, g\rangle = \frac{1}{[2\pi]^n} \int\limits_{\|\boldsymbol{\xi}\|<R} \hat{f}(\boldsymbol{\xi})\overline{\hat{g}(\boldsymbol{\xi})} d\boldsymbol{\xi}. \tag{4.89}$$

The proof of the lemma is a consequence of the Parseval formula, it is left as an exercise for the reader. Using (4.89) gives

$$\|f_R - f\|_{L^2}^2 = \frac{1}{[2\pi]^n} \int\limits_{\|\boldsymbol{\xi}\|\geq R} |\hat{f}(\boldsymbol{\xi})|^2 d\boldsymbol{\xi}.$$

This implies that $\underset{R\to\infty}{LIM} F_R = f$. □

Remark 4.5.2. The extension of the Fourier transform to functions in $L^2(\mathbb{R}^n)$ has many nice properties. In particular the range of the Fourier transform on L^2 is exactly $L^2(\mathbb{R}^n)$. However the formula for the Fourier transform as an integral is *purely symbolic*. The Fourier transform itself is only defined as a LIM; for a given $\boldsymbol{\xi}$ the pointwise limit

$$\lim_{R\to\infty} \int\limits_{\|\mathbf{x}\|<R} f(\mathbf{x})^{-i\mathbf{x}\cdot\boldsymbol{\xi}} d\mathbf{x}$$

may or may not exist.

We conclude this section with an enumeration of the elementary properties of the Fourier transform for functions of n-variables. As before these hold for integrable or square-integable functions and follow from elementary properties of the integral.

1. Linearity:

   The Fourier transform is a linear operation, if $\alpha \in \mathbb{C}$ then

   $$\mathcal{F}(f+g) = \mathcal{F}(f) + \mathcal{F}(g), \quad \mathcal{F}(\alpha f) = \alpha\mathcal{F}(f).$$

2. Scaling:

   The Fourier transform of $f(ax)$, a function dilated by $a \in \mathbb{R}$ is given by

   $$\begin{aligned} \int\limits_{\mathbb{R}^n} f(ax)e^{-i\boldsymbol{\xi}\cdot\mathbf{x}} d\mathbf{x} &= \int\limits_{\mathbb{R}^n} f(\mathbf{y})e^{-i\frac{\boldsymbol{\xi}\cdot\mathbf{y}}{a}} \frac{d\mathbf{y}}{a^n} \\ &= \frac{1}{a^n}\hat{f}(\frac{\boldsymbol{\xi}}{a}). \end{aligned} \tag{4.90}$$

3. Translation:

   Let $f_{\mathbf{t}}$ be the function f shifted by the vector $\mathbf{t}$, $f_{\mathbf{t}}(\mathbf{x}) = f(\mathbf{x}-\mathbf{t})$. The Fourier transform of $f_{\mathbf{t}}(\boldsymbol{\xi})$ is given by

   $$\begin{aligned} \widehat{f_{\mathbf{t}}}(\boldsymbol{\xi}) &= \int\limits_{\mathbb{R}^n} f(\mathbf{x}-\mathbf{t})e^{-i\boldsymbol{\xi}\cdot\mathbf{x}} d\mathbf{x} \\ &= \int\limits_{\mathbb{R}^n} f(\mathbf{y})e^{-i\boldsymbol{\xi}\cdot(\mathbf{y}+\mathbf{t})} d\mathbf{y} \\ &= e^{-i\boldsymbol{\xi}\cdot\mathbf{t}}\hat{f}(\boldsymbol{\xi}). \end{aligned} \tag{4.91}$$

4. Reality:
   If $f(\mathbf{x})$ is real valued then $\hat{f}(\boldsymbol{\xi}) = \overline{\hat{f}(-\boldsymbol{\xi})}$.

5. Evenness:
   If f is even then $\hat{f}$ is real valued, if f is odd then $\hat{f}$ is purely imaginary valued.

Exercises

Exercise 4.5.10. Give the details of the proof of the n-dimensional Parseval formula.

Exercise 4.5.11. Show that (4.87) implies that

$$\int_{\mathbb{R}^n} f(\mathbf{x})\overline{g(\mathbf{x})}d\mathbf{x} = \int_{\mathbb{R}^n} \hat{f}(\boldsymbol{\xi})\overline{\hat{g}(\boldsymbol{\xi})}\frac{d\boldsymbol{\xi}}{[2\pi]^n}.$$

Exercise 4.5.12. Prove Lemma 4.5.1.

Exercise 4.5.13. Verify properties (4) and (5).

Exercise 4.5.14. Prove that the Fourier transform of a radial function is also a radial function and formula (4.92).

4.5.4 The Fourier transform on radial functions

Recall that a function which only depends on $\|\mathbf{x}\|$ is said to be *radial*. The Fourier transform of a radial function is also radial and can be given by a 1-dimensional integral transform.

Theorem 4.5.4. *Suppose that $f(\mathbf{x}) = F(\|\mathbf{x}\|)$ is an integrable function, then the Fourier transform of f is given by the one-dimensional integral transform*

$$\hat{f}(\boldsymbol{\xi}) = \frac{c_n}{\|\boldsymbol{\xi}\|^{\frac{n-2}{2}}} \int_0^\infty J_{\frac{n-2}{2}}(r\|\boldsymbol{\xi}\|)F(r)r^{\frac{n}{2}}dr. \tag{4.92}$$

Here c_n is a constant depending on the dimension.

If $\mathrm{Re}(\nu) > -\frac{1}{2}$ then $J_\nu(z)$, the order ν Bessel function is defined by the integral

$$J_\nu(z) = \frac{\left(\frac{z}{2}\right)^\nu}{\Gamma\left(\nu+\frac{1}{2}\right)\Gamma\left(\frac{1}{2}\right)} \int_0^\pi e^{iz\cos(\theta)}\sin^{2\nu}(\theta)d\theta.$$

Proof. The derivation of (4.92) uses polar coordinates on $\mathbb{R}^n$. Let $\mathbf{x} = r\boldsymbol{\omega}$ where r is a non-negative number and $\boldsymbol{\omega}$ belongs to the unit $(n-1)$-sphere, S^{n-1}. In these coordinates the volume form on $\mathbb{R}^n$ is

$$d\mathbf{x} = r^{n-1}drdV_{S^{n-1}},$$

here $dV_{S^{n-1}}$ is the volume form on S^{n-1}. In polar coordinates, the Fourier transform of f is given by

$$\hat{f}(\boldsymbol{\xi}) = \int_0^\infty \int_{S^{n-1}} F(r)e^{-ir\langle\boldsymbol{\omega},\boldsymbol{\xi}\rangle}dV_{S^{n-1}}r^{n-1}dr. \tag{4.93}$$

It not difficult to show that the integral over S^{n-1} only depends on $\|\boldsymbol{\xi}\|$ and therefore it suffices to evaluate it for $\boldsymbol{\xi} = (0, \ldots, 0, \|\boldsymbol{\xi}\|)$. Points on the $(n-1)$-sphere can be expressed in the form

$$\boldsymbol{\omega} = \sin\theta(\boldsymbol{\omega}', 0) + (0, \ldots, 0, \cos\theta),$$

where $\boldsymbol{\omega}'$ is a point on the unit $(n-2)$-sphere and $\theta \in [0, \pi]$. Using this parametrization for S^{n-1} we obtain a formula for the volume form,

$$dV_{S^{n-1}} = \sin^{n-2}\theta dV_{S^{n-2}}. \tag{4.94}$$

Using these observations, the spherical integral in (4.93) becomes

$$\begin{aligned}\int\limits_{S^{n-1}} e^{-ir\langle\boldsymbol{\omega},\boldsymbol{\xi}\rangle} dV_{S^{n-1}} &= \int\limits_0^\pi \int\limits_{S^{n-2}} e^{-ir\|\boldsymbol{\xi}\|\cos\theta} \sin^{n-2}\theta dV_{S^{n-2}} d\theta \\ &= \sigma_{n-2} \int\limits_0^\pi e^{-ir\|\boldsymbol{\xi}\|\cos\theta} \sin^{n-2}\theta d\theta.\end{aligned} \tag{4.95}$$

The coefficient σ_{n-2} is the $(n-2)$-dimensional volume of S^{n-2}. Comparing this integral with the definition of the Bessel function gives (4.92). □

Example 4.5.3. The Fourier transform of the characteristic function of the unit ball $B_1 \subset \mathbb{R}^n$ is given by the radial integral

$$\widehat{\chi_{B_1}}(\boldsymbol{\xi}) = \frac{c_n}{\|\boldsymbol{\xi}\|^{\frac{n-2}{2}}} \int\limits_0^1 J_{\frac{n-2}{2}}(r\|\boldsymbol{\xi}\|) r^{\frac{n}{2}} dr.$$

Using formula 6.561.5 in [20] gives

$$\widehat{\chi_{B_1}}(\boldsymbol{\xi}) = \frac{c_n}{\|\boldsymbol{\xi}\|^{\frac{n}{2}}} J_{\frac{n}{2}}(\|\boldsymbol{\xi}\|).$$

As $\|\boldsymbol{\xi}\|$ tends to infinity the Bessel function is a oscillatory term times $[\sqrt{\|\boldsymbol{\xi}\|}]^{-1}$. Overall we have the estimate

$$\widehat{\chi_{B_1}}(\boldsymbol{\xi}) \leq \frac{C}{(1 + \|\boldsymbol{\xi}\|)^{\frac{n+1}{2}}}.$$

Exercises

Exercise 4.5.15. Prove that the spherical integral in (4.93) only depends on $\|\boldsymbol{\xi}\|$.

Exercise 4.5.16. Verify the parametrization of the $(n-1)$-sphere used to obtain (4.94) as well as this formula.

Exercise 4.5.17. Determine the constant c_n in (4.92).

Exercise 4.5.18. Using (4.94) show that σ_n the n-volume of S^n is given by

$$\sigma_n = \frac{2\pi^{\frac{n+1}{2}}}{\Gamma\left(\frac{n+1}{2}\right)}.$$

Exercise 4.5.19. Using the connection between the one-dimensional integral transform defined in (4.92) and the n-dimensional Fourier transform, find a formula for the inverse of this transform. Hint: Use symmetry, this does not require any computation!

4.5.5 The failure of localization in higher dimensions*

The localization principle is a remarkable feature of the 1-dimensional Fourier transform. Suppose that f is an integrable function defined on $\mathbb{R}$. According to the localization principle the convergence of the partial inverse

$$f_R(x) = \frac{1}{2\pi} \int\limits_{-R}^{R} \hat{f}(\xi) e^{ix\xi} d\xi$$

to $f(x)$ only depends on the behavior of f in an interval about x. This is a uniquely one dimensional phenomenon. In this section we give an example due to Pinsky showing the failure of the localization principle in three dimensions. A complete discussion of this phenomenon can be found in [57].

Pinsky's example is very simple, it concerns $f(\mathbf{x}) = \chi_{B_1}(\mathbf{x})$, the characteristic function of the unit ball. The Fourier transform of f was computed in example 4.5.3, it is

$$\hat{f}(\boldsymbol{\xi}) = \frac{cJ_{\frac{3}{2}}(\|\boldsymbol{\xi}\|)}{\|\boldsymbol{\xi}\|^{\frac{3}{2}}}.$$

In this example c denotes various positive constants. Using formula 8.464.3 in [20] this can be re-expressed in terms of elementary functions by

$$\hat{f}(\boldsymbol{\xi}) = \frac{c[\|\boldsymbol{\xi}\| \cos(\|\boldsymbol{\xi}\|) - \sin(\|\boldsymbol{\xi}\|)]}{\|\boldsymbol{\xi}\|^3}.$$

Using polar coordinates, we compute the partial inverse:

$$\begin{aligned} f_R(0) &= \frac{c}{[2\pi]^3} \int\limits_0^R \left[\cos(r) - \frac{\sin(r)}{r}\right] dr \\ &= c\left[\sin(R) - \int\limits_0^R \frac{\sin(r)}{r} dr\right]. \end{aligned} \tag{4.96}$$

The last integral has a limit as $R \to \infty$ however $\sin(R)$ does not! Thus $f_R(0)$ remains bounded as R tends to infinity but does not converge.

Remark 4.5.3. The reader interested in complete proofs for the results in this section as well as further material is directed to [40], for the one dimensional case or [72], for higher dimensions.

Exercises

Exercise 4.5.20. Prove the existence of the limit

$$\lim_{R\to\infty} \int\limits_0^R \frac{\sin(r)dr}{r}.$$

Chapter 5

Convolution

In the previous chapter we introduced the Fourier transform with two purposes in mind: (1) Finding the inverse for the Radon transform. (2) Applying it to signal and image processing problems. Indeed (1) is a special case of (2). In this chapter we introduce a fundamental operation, called the *convolution product.* Convolution is intimately connected to the Fourier transform. Because there is a very efficient algorithm for approximating the Fourier transform and its inverse, convolution lies at the heart of many practical filters. After defining the convolution product for functions defined on $\mathbb{R}^n$ and establishing its basic properties we briefly turn our attention to filtering theory.

5.1 Convolution

See: A.7.1.

For applications to medical imaging we use convolution in 1-, 2- and 3-dimensions. The definition and formal properties of this operation do not depend on the dimension and we therefore define it and consider its properties for functions defined on $\mathbb{R}^n$.

Definition 5.1.1. If f is an integrable function defined on $\mathbb{R}^n$ and g is bounded then the *convolution product* of f and g is defined by the absolutely convergent integral

$$f * g(\mathbf{x}) = \int\limits_{\mathbb{R}^n} f(\mathbf{x}-\mathbf{y})g(\mathbf{y})d\mathbf{y}. \tag{5.1}$$

Remark 5.1.1. There are many different conditions under which this operation is defined. If the product $f(\mathbf{y})g(\mathbf{x}-\mathbf{y})$ is an integrable function of $\mathbf{y}$ then $f * g(\mathbf{x})$ is defined by an absolutely convergent integral. For example, if g is bounded with bounded support then it is only necessary that f be locally integrable in order for $f*g$ to be defined. In this chapter we use functional analytic methods to extend the definition of convolution to situations where these integals are not absolutely convergent. This closely follows the pattern established to extend the Fourier transform to $L^2(\mathbb{R}^n)$.

Convolution provides a general framework for analyzing moving averages. To better understand it, first think of f as an input "signal" and g as a non-negative weight function, then $f * g(\mathbf{x})$ is a weighted average of the values of f. The contribution of $f(\mathbf{x}-\mathbf{y})$ to $f * g(\mathbf{x})$ is given weight $g(\mathbf{y})$.

Example 5.1.1. Let $g(x) = \frac{1}{2}\chi_{[-1,1]}(x)$; if f is a locally integrable function then the integrals in (5.1) make sense. The convolution $f * g$ at x is

$$f * g(x) = \frac{1}{2}\int\limits_{x-1}^{x+1} f(y)dy.$$

This is the ordinary average of f over the interval $[x-1, x+1]$. If $f = \chi_{[-1,1]}$ then

$$f * g(x) = \begin{cases} 0 & \text{if } |x| > 2, \\ \frac{|2+x|}{2} & \text{if } -2 \leq x \leq 0, \\ \frac{|2-x|}{2} & \text{if } 0 \leq x \leq 2. \end{cases}$$

Example 5.1.2. Let $g(\mathbf{x}) = c_n r^{-n}\chi_{B_r}(\|\mathbf{x}\|)$; here B_r is the ball of radius r in $\mathbb{R}^n$ and c_n^{-1} is the volume of B_1. Again $f * g$ is defined for any locally integrable function and is given by

$$\begin{aligned} f * g(\mathbf{x}) &= \int\limits_{\mathbb{R}^n} f(\mathbf{x}-\mathbf{y})g(\mathbf{y})d\mathbf{y} \\ &= c_n r^{-1}\int\limits_{B_r} f(\mathbf{x}-\mathbf{y})d\mathbf{y}. \end{aligned} \tag{5.2}$$

It is the average of the values of f over points in $B_r(\mathbf{x})$.

Convolution also appears in the partial inverse of the Fourier transform. In this case the weighting function assumes both positive and negative values.

Example 5.1.3. Let f belong to either $L^1(\mathbb{R})$ or $L^2(\mathbb{R})$. In section 4.4.1 we defined the partial inverse of the Fourier transform

$$S_R(f)(x) = \frac{1}{2\pi}\int\limits_{-R}^{R} \hat{f}(\xi)e^{ix\xi}d\xi. \tag{5.3}$$

This can be represented as a convolution,

$$S_R(f) = f * D_R,$$

where

$$D_R(x) = \frac{R\,\text{sinc}(Rx)}{\pi}.$$

For functions in either space this convolution is given by an absolutely convergent integral.

The convolution product satisfies many estimates, the simplest is a consequence of the triangle inequality for integrals:

$$\|f * g\|_\infty \leq \|f\|_{L^1}\|g\|_\infty. \tag{5.4}$$

We now establish another estimate which, via Theorem 4.2.3, extends the domain of the convolution product.

Proposition 5.1.1. *Suppose that f and g are integrable and g is bounded then $f * g$ is absolutely integrable and*

$$\|f * g\|_{L^1} \leq \|f\|_{L^1}\|g\|_{L^1}. \tag{5.5}$$

Proof. It follows from the triangle inequality that

$$\begin{aligned}\int\limits_{\mathbb{R}^n} |f * g(\mathbf{x})|d\mathbf{x} &\leq \int\limits_{\mathbb{R}^n}\int\limits_{\mathbb{R}^n} |f(\mathbf{x}-\mathbf{y})g(\mathbf{y})|d\mathbf{y}d\mathbf{x} \\ &= \int\limits_{\mathbb{R}^n}\int\limits_{\mathbb{R}^n} |f(\mathbf{x}-\mathbf{y})g(\mathbf{y})|d\mathbf{x}d\mathbf{y}.\end{aligned} \tag{5.6}$$

Going from the first to the second lines we interchanged the order of the integrations. This is allowed by Fubini's theorem, since $f(\mathbf{y})g(\mathbf{x}-\mathbf{y})$ is absolutely integable over $\mathbb{R}^n \times \mathbb{R}^n$. We change variables in the $\mathbf{x}$-integral by setting $\mathbf{t} = \mathbf{x}-\mathbf{y}$ to get

$$\|f * g\|_{L^1} \leq \int\limits_{\mathbb{R}^n}\int\limits_{\mathbb{R}^n} |f(\mathbf{t})||g(\mathbf{y})|d\mathbf{t}d\mathbf{y} = \|f\|_{L^1}\|g\|_{L^1}.$$

□

For a fixed f in $L^1(\mathbb{R}^n)$ the map from bounded, integrable functions to $L^1(\mathbb{R}^n)$ defined by $C_f(g) = f * g$ is linear and satisfies (5.5). As bounded functions are dense in $L^1(\mathbb{R}^n)$, Theorem 4.2.3 applies to show that C_f extends to define a map from $L^1(\mathbb{R}^n)$ to itself. Because f is an arbitrary integrable function, convolution extends as a map from $L^1(\mathbb{R}^n) \times L^1(\mathbb{R}^n)$ to $L^1(\mathbb{R}^n)$. The following proposition summarizes these observations.

Proposition 5.1.2. *The convolution product extends to define a continuous map from $L^1(\mathbb{R}^n) \times L^1(\mathbb{R}^n)$ to $L^1(\mathbb{R}^n)$ which satisfies* (5.5)*.*

Remark 5.1.2. If f and g are both in $L^1(\mathbb{R}^n)$ then the integral defining $f * g(\mathbf{x})$ may not converge for every $\mathbf{x}$. The fact that $f(\mathbf{y})g(\mathbf{x}-\mathbf{y})$ is integrable over $\mathbb{R}^n \times \mathbb{R}^n$ implies that

$$\int\limits_{\mathbb{R}^n} f(\mathbf{y})g(\mathbf{x}-\mathbf{y})d\mathbf{y}$$

might diverge for $\mathbf{x}$ belonging to a set of measure zero. An inequality analogous to (5.5) holds for any $1 \leq p \leq \infty$. That is, if $f \in L^p(\mathbb{R}^n)$ and $g \in L^1(\mathbb{R}^n)$ then $f * g$ is defined as an element of $L^p(\mathbb{R}^n)$, satisfying the estimate

$$\|f * g\|_{L^p} \leq \|f\|_{L^p}\|g\|_{L^1}. \tag{5.7}$$

The proof of this statement is left to the exercises.

Example 5.1.4. Some decay conditions are required for $f * g$ to be defined. If $f(x) = [\sqrt{1+|x|}]^{-1}$ then

$$f * f(x) = \int\limits_{-\infty}^{\infty} \frac{1}{\sqrt{1+|y|}}\frac{1}{\sqrt{1+|x-y|}}dy = \infty \text{ for all } x.$$

If, on the other hand, we let $g = [\sqrt{1+|x|}]^{-(1+\epsilon)}$, for any positive ϵ then $f * g$ is defined.

The basic properties of integration lead to certain algebraic properties for the convolution product.

Proposition 5.1.3. *Suppose that f_1, f_2, f_3 belong to $L^1(\mathbb{R}^n)$ then the following identities hold:*

$$\begin{aligned} f_1 * f_2 &= f_2 * f_1, \\ f_1 * (f_2 + f_3) &= f_1 * f_2 + f_1 * f_3, \\ f_1 * (f_2 * f_3) &= (f_1 * f_2) * f_3. \end{aligned} \tag{5.8}$$

Proof. We prove the first assertion; it suffices to assume that f_2 is bounded, the general case then follows from (5.5). The definition states that

$$f_1 * f_2(\mathbf{x}) = \int\limits_{\mathbb{R}^n} f_1(\mathbf{y}) f_2(\mathbf{x}-\mathbf{y}) d\mathbf{y}.$$

Letting $\mathbf{t} = \mathbf{x} - \mathbf{y}$ this integral becomes

$$\int\limits_{\mathbb{R}^n} f_1(\mathbf{x}-\mathbf{t}) f_2(\mathbf{t}) d\mathbf{t} = f_2 * f_1(\mathbf{x}).$$

The proofs of the remaining parts are left to the exercises. □

Taken together, these identities say that convolution defines a multiplication on $L^1(\mathbb{R}^n)$ which is commutative, distributive and associative. The only thing missing is a multiplicative unit, that is a function $i \in L^1(\mathbb{R}^n)$ so that $f * i = f$ for every f in $L^1(\mathbb{R}^n)$. It is not hard to convince ourselves that such a function cannot exist. For if

$$f(\mathbf{x}) = \int\limits_{\mathbb{R}^n} f(\mathbf{x}-\mathbf{y}) i(\mathbf{y}) d\mathbf{y},$$

for every point $\mathbf{x}$ and every function $f \in L^1(\mathbb{R}^n)$ then i must vanish for $\mathbf{x} \neq 0$. But in this case $f * i \equiv 0$ for any function $f \in L^1(\mathbb{R}^n)$. In section 5.3 we return to this point.

A reason that the convolution product is so important in applications is that the Fourier transform converts convolution into ordinary pointwise multiplication.

Theorem 5.1.1. *Suppose that f and g are absolutely integrable then*

$$\mathcal{F}(f * g) = \mathcal{F}(f)\mathcal{F}(g). \tag{5.9}$$

Proof. The convolution, $f * g$ is absolutely integrable and therefore has a Fourier transform. Because $f(\mathbf{x}-\mathbf{y})g(\mathbf{y})$ is an absolutely integrable function of $(\mathbf{x}, \mathbf{y})$, the following manipulations are easily justified,

$$\begin{aligned} \mathcal{F}(f * g)(\boldsymbol{\xi}) &= \int\limits_{\mathbb{R}^n} (f * g)(\mathbf{x}) e^{-i\langle \boldsymbol{\xi}, \mathbf{x}\rangle} d\mathbf{x} \\ &= \int\limits_{\mathbb{R}^n} \int\limits_{\mathbb{R}^n} f(\mathbf{x}-\mathbf{y}) g\mathbf{y}) e^{-i\langle \boldsymbol{\xi}, \mathbf{x}\rangle} d\mathbf{y} d\mathbf{x} \\ &= \int\limits_{\mathbb{R}^n} \int\limits_{\mathbb{R}^n} f(\mathbf{t}) g(\mathbf{y}) e^{-i\langle \boldsymbol{\xi}, (\mathbf{y}+\mathbf{t})\rangle} d\mathbf{t} d\mathbf{y} \\ &= \hat{f}(\boldsymbol{\xi}) \hat{g}(\boldsymbol{\xi}). \end{aligned} \tag{5.10}$$

□

Remark 5.1.3. The conclusion of Theorem 5.1.1 remains true if $f \in L^2(\mathbb{R}^n)$ and $g \in L^1(\mathbb{R}^n)$. In this case $f * g$ also belongs to $L^2(\mathbb{R}^n)$. Note that $\hat{g}$ is a bounded function, so that $\hat{f}\hat{g}$ belongs to $L^2(\mathbb{R}^n)$ as well.

Example 5.1.5. Let $f = \chi_{[-1,1]}$. Formula (5.9) simplifies the computation of the Fourier transform for $f * f$ or even the j-fold convolution of f with itself

$$f *_j f \stackrel{d}{=} f \underbrace{* \cdots *}_{j-\text{times}} f.$$

In this case

$$\mathcal{F}(f *_j f)(\xi) = [2\,\text{sinc}(\xi)]^j.$$

Example 5.1.6. A partial inverse for the Fourier transform in n-dimensions is defined by

$$S_R^n(f) = \frac{1}{[2\pi]^n} \int_{-R}^{R} \cdots \int_{-R}^{R} \hat{f}(\boldsymbol{\xi}) e^{i\langle \mathbf{x}, \boldsymbol{\xi} \rangle} d\boldsymbol{\xi}.$$

The Fourier transform of the function

$$D_R^n(\mathbf{x}) = \left[\frac{R}{\pi}\right]^n \prod_{j=1}^{n} \text{sinc}(Rx_j)$$

is $\chi_{[-R,R]}(\xi_1) \cdots \chi_{[-R,R]}(\xi_n)$ and therefore Theorem 5.1.1 implies that

$$S_R^n(f) = D_R^n * f.$$

Exercises

Exercise 5.1.1. For $f \in L^1(\mathbb{R})$ define

$$f_B(x) = \begin{cases} f(x) & \text{if } |f(x)| \leq B, \\ 0 & \text{if } |f(x)| > B. \end{cases}$$

Show that $\lim_{B \to \infty} \|f - f_B\|_{L^1} = 0$. Use this fact and the inequality, (5.5) to show that the sequence $< f_B * g >$ has a limit in $L^1(\mathbb{R})$.

Exercise 5.1.2. Prove the remaining parts of Proposition 5.1.3. Explain why it suffices to prove these identities for *bounded* integrable functions.

Exercise 5.1.3. Compute $\chi_{[-1,1]} *_j \chi_{[-1,1]}$ for $j = 2, 3, 4$ and plot these functions on a single graph.

Exercise 5.1.4. Prove that $\|f * g\|_{L^2} \leq \|f\|_{L^2} \|g\|_{L^1}$. Hint: Use the Cauchy-Schwarz inequality.

Exercise 5.1.5. * For $1 < p < \infty$ use Hölder's inequality to show that $\|f * g\|_{L^p} \leq \|f\|_{L^p} \|g\|_{L^1}$.

Exercise 5.1.6. Show that $\mathcal{F}(D_R^n)(\boldsymbol{\xi}) = \chi_{[-R,R]}(\xi_1) \cdots \chi_{[-R,R]}(\xi_n)$.

Exercise 5.1.7. Prove that the conclusion of Theorem 5.1.1 remains true if $f \in L^2(\mathbb{R}^n)$ and $g \in L^1(\mathbb{R}^n)$. Hint: Use the estimate $\|f * g\|_{L^2} \leq \|f\|_{L^2}\|g\|_{L^1}$ to reduce to a simpler case.

Exercise 5.1.8. Show that there does not exist an integrable function i so that $i*f = f$ for every integrable function f. Hint: Use Theorem 5.1.1 and the Riemann-Lebesgue Lemma.

Exercise 5.1.9. A different partial inverse for the n-dimensional Fourier transform is defined by

$$\Sigma_R(f) = \frac{1}{[2\pi]^n} \int\limits_{\|\boldsymbol{\xi}\| \leq R} \hat{f}(\boldsymbol{\xi}) e^{i\langle \mathbf{x}, \boldsymbol{\xi}\rangle} d\boldsymbol{\xi}.$$

This can also be expressed as the convolution of f with a function F_R^n. Find an explicit formula for F_R^n.

5.1.1 Shift invariant filters

In engineering essentially any operation which maps inputs to outputs is called a *filter*. Since most inputs and outputs are represented by functions, a filter is usually a map from one space of functions to another. The filter is a *linear filter* if this map of function spaces is linear. In practice many filtering operations are given by convolution with a fixed function. If $\psi \in L^1(\mathbb{R}^n)$ then

$$C_\psi(g) = \psi * g,$$

defines such a filter. A filter which takes bounded inputs to bounded outputs is called a *stable filter*. The estimate (5.4) shows that any filter defined by convolution with an L^1-function is stable. Indeed the estimate (5.7) shows that such filters act continuously on many function spaces.

Filters defined by convolution have an important physical property: they are *shift invariant*.

Definition 5.1.2. For $\boldsymbol{\tau} \in \mathbb{R}^n$ the *shift of f by $\boldsymbol{\tau}$* is the function $f_{\boldsymbol{\tau}}$, defined by

$$f_{\boldsymbol{\tau}}(\mathbf{x}) = f(\mathbf{x} - \boldsymbol{\tau}).$$

A filter, $\mathcal{A}$ mapping functions defined on $\mathbb{R}^n$ to functions defined on $\mathbb{R}^n$ is *shift invariant* if

$$\mathcal{A}(f_{\boldsymbol{\tau}}) = (\mathcal{A}f)_{\boldsymbol{\tau}}.$$

If $n = 1$ and the input is a function of time, then a filter is shift invariant if the action of the filter does not depend on *when* the input arrives. If the input is a function of spatial variables, then a filter is shift invariant if its action does not depend on *where* the input is located.

Proposition 5.1.4. *A filter defined by convolution is shift invariant.*

Proof. The proof is a simple change of variables.

$$\begin{aligned} C_\psi(f_{\boldsymbol{\tau}})(\mathbf{x}) &= \int\limits_{\mathbb{R}^n} \psi(\mathbf{x} - \mathbf{y}) f(\mathbf{y} - \boldsymbol{\tau}) d\mathbf{y} \\ &= \int\limits_{\mathbb{R}^n} \psi(\mathbf{x} - \boldsymbol{\tau} - \mathbf{w}) f(\mathbf{w}) d\mathbf{w} \\ &= C_\psi(f)(\mathbf{x} - \boldsymbol{\tau}). \end{aligned} \tag{5.11}$$

In going from the first to the second line we used the change of variable $\mathbf{w} = \mathbf{y} - \boldsymbol{\tau}$. ∎

In a certain sense the converse is also true: "Any" shift invariant, linear filter can be represented by convolution. What makes this a little complicated is that the function ψ may need to be replaced by a generalized function.

Beyond the evident simplicity of shift invariance, this class of filters is important for another reason: Theorem 5.1.1 shows that the output of such a filter can be computed using the Fourier transform and its inverse, explicitly

$$C_\psi(f) = \mathcal{F}^{-1}(\hat{\psi}\hat{f}). \tag{5.12}$$

This is significant because the Fourier transform has a very efficient, approximate numerical implementation.

Example 5.1.7. Let $\psi = \frac{1}{2}\chi_{[-1,1]}$, the convolution $\psi * f$ is the moving average of f over intervals of length 2. It can be computed using the Fourier transform by,

$$\psi * f(x) = \frac{1}{2\pi}\int\limits_{-\infty}^{\infty} \operatorname{sinc}(\xi)\hat{f}(\xi)e^{ix\xi}d\xi.$$

Exercises

Exercise 5.1.10. For each of the following filters, decide if is shift invariant or non-shift invariant.

(1). Translation: $\mathcal{A}_{\boldsymbol{\tau}}(f)(\mathbf{x}) \overset{d}{=} f(\mathbf{x}-\boldsymbol{\tau})$.

(2). Scaling: $\mathcal{A}_\epsilon(f)(\mathbf{x}) \overset{d}{=} \frac{1}{\epsilon^n}f\left(\frac{\mathbf{x}}{\epsilon}\right)$.

(3). Multiplication by a function: $\mathcal{A}_\psi(f) \overset{d}{=} \psi f$.

(4). Indefinite integral from 0: $\mathcal{I}_0(f)(x) \overset{d}{=} \int_0^x f(y)dy$.

(5). Indefinite integral from $-\infty$: $\mathcal{I}_{-\infty}(f)(x) \overset{d}{=} \int_{-\infty}^x f(y)dy$.

(6). Time reversal: $\mathcal{T}_r(f)(x) \overset{d}{=} f(-x)$.

(7). Integral filter: $f \mapsto \int_{-\infty}^{\infty} xyf(y)dy$.

(8). Differentiation: $\mathcal{D}(f)(x) = f'(x)$.

Exercise 5.1.11. Suppose that $\mathcal{A}$ and $\mathcal{B}$ are shift invariant. Show that their composition $\mathcal{A}\circ\mathcal{B}(f) \overset{d}{=} \mathcal{A}(\mathcal{B}(f))$ is also shift invariant.

5.2 Convolution and regularity

Generally speaking the averages of a function are smoother than the function itself. If f is a locally integrable function and g is continuous, with bounded support then $f * g$ is continuous. Let $\boldsymbol{\tau}$ be a vector in $\mathbb{R}^n$ then

$$\lim_{\boldsymbol{\tau}\to\mathbf{0}}[f*g(\mathbf{x}+\boldsymbol{\tau}) - f*g(\mathbf{x})] = \lim_{\boldsymbol{\tau}\to\mathbf{0}}\int\limits_{\mathbb{R}^n} f(\mathbf{y})[g(\mathbf{x}+\boldsymbol{\tau}-\mathbf{y}) - g(\mathbf{x}-\mathbf{y})]d\mathbf{y}.$$

Because g has bounded support it follows that the limit on the right can be taken inside the integral, showing that

$$\lim_{\boldsymbol{\tau}\to\mathbf{0}} f*g(\mathbf{x}+\boldsymbol{\tau}) = f*g(\mathbf{x}).$$

This argument can be repeated with difference quotients to prove the following result.

Proposition 5.2.1. *Suppose that f is locally integrable, g has bounded support and k continuous derivatives, then $f*g$ also has k continuous derivatives. For any multi-index $\boldsymbol{\alpha}$ with $|\boldsymbol{\alpha}|\leq k$ we have*

$$\partial_{\mathbf{x}}^{\boldsymbol{\alpha}}(f*g) = f*(\partial_{\mathbf{x}}^{\boldsymbol{\alpha}} g). \tag{5.13}$$

Remark 5.2.1. This result is also reasonable from the point of view of the Fourier transform. Suppose that g has k integrable derivatives, then Theorem 4.5.2 shows that

$$|\hat{g}(\boldsymbol{\xi})| \leq \frac{C}{(1+\|\boldsymbol{\xi}\|)^k}.$$

If f is either integrable or square-integrable then the Fourier transform of $f*g$ satisfies an estimate of the form

$$|\mathcal{F}(f*g)(\boldsymbol{\xi})| \leq \frac{C|\hat{f}(\boldsymbol{\xi})|}{(1+\|\boldsymbol{\xi}\|)^k}.$$

This shows that the Fourier transform of $f*g$ has a definite improvement in its rate of decay over that of f and therefore $f*g$ is commensurately smoother.

5.2.1 Approximation by smooth functions

If φ is a smooth function with bounded support and f is locally integrable then $\varphi*f$ is a smooth function. Hence convolution gives a method for approximating integrable functions by smooth functions. To do that we choose a non-negative, even, infinitely differentiable function, ϕ supported in $[-1,1]$. An example is given by the function

$$\phi(t) = \begin{cases} e^{-\frac{1}{1-t^2}} & \text{if } |t|<1, \\ 0 & \text{if } |t|\geq 1. \end{cases}$$

Let $\varphi(\mathbf{x}) = c_n\phi(\|\mathbf{x}\|)$, with the constant c_n selected so that

$$\int_{\mathbb{R}^n} \varphi(\mathbf{x})d\mathbf{x} = 1.$$

For $\epsilon>0$ let

$$\varphi_\epsilon(\mathbf{x}) = \epsilon^{-n}\varphi(\frac{\mathbf{x}}{\epsilon}), \tag{5.14}$$

see figure 5.1. Observe that φ_ϵ is supported in the ball of radius ϵ. Using the change of variables $\epsilon\mathbf{y}=\mathbf{x}$ gives

$$\int_{\mathbb{R}^n} \varphi_\epsilon(\mathbf{x})d\mathbf{x} = \int_{\mathbb{R}^n} \frac{1}{\epsilon^n}\varphi(\frac{\mathbf{x}}{\epsilon})d\mathbf{x} = \int_{\mathbb{R}^n} \varphi(\mathbf{y})d\mathbf{y} = 1. \tag{5.15}$$

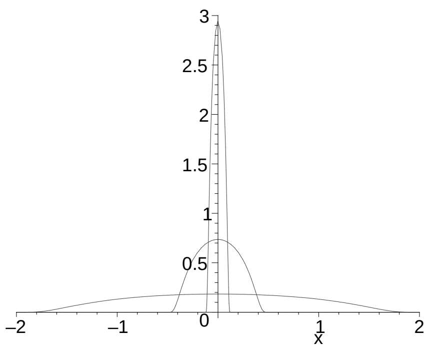


Figure 5.1: Graphs of φ_ϵ, with $\epsilon = .5, 2, 8$.

This identity allows the difference between f and $\varphi_\epsilon * f$ to be expressed in a convenient form

$$\varphi_\epsilon * f(\mathbf{x}) - f(\mathbf{x}) = \int_{B_\epsilon(\mathbf{x})} [f(\mathbf{y}) - f(\mathbf{x})]\varphi_\epsilon(\mathbf{x}-\mathbf{y})d\mathbf{y}. \tag{5.16}$$

The integral is over the ball of radius ϵ, centered at $\mathbf{x}$. It is therefore reasonable to expect that, as ϵ goes to zero, $\varphi_\epsilon * f$ converges, in some sense, to f. Note that

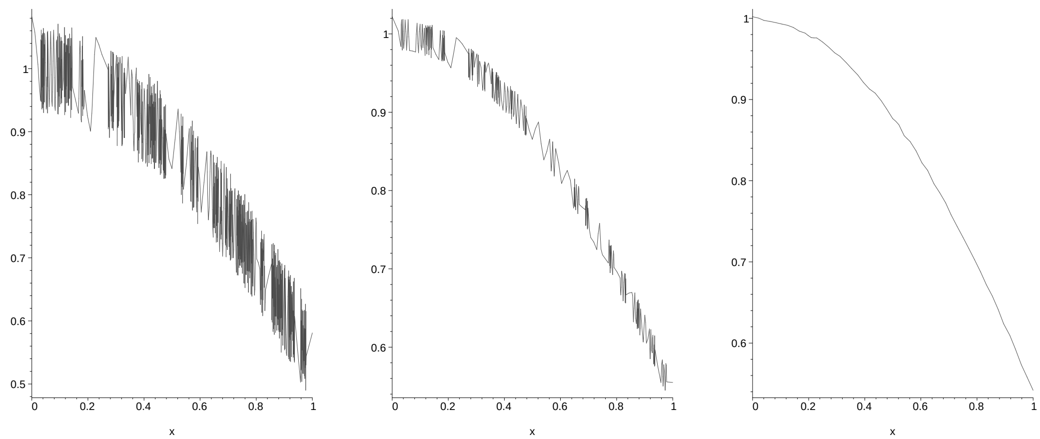
$$\hat{\varphi}(0) = \int_{-\infty}^{\infty} \varphi(x)dx = 1.$$

Heuristically this gives another way to understand what happens to $\varphi_\epsilon * f$ as $\epsilon \to 0$:

$$\begin{aligned}\widehat{\varphi_\epsilon * f}(\xi) &= \hat{\varphi}_\epsilon(\xi)\hat{f}(\xi) \\ &= \hat{\varphi}(\epsilon\xi)\hat{f}(\xi) \to \hat{\varphi}(0)\hat{f}(\xi) \\ &= \hat{f}(\xi) \quad \text{as} \quad \epsilon \to 0.\end{aligned} \tag{5.17}$$

Again, in some sense, $\varphi_\epsilon * f$ converges to f as ϵ tends to 0.

If we think of f as representing a noisy signal then $\varphi_\epsilon * f$ is a smoothed out version of f. In applications ϵ is a measure of the resolution available in $\varphi_\epsilon * f$. A larger ϵ results in a more blurred, but less noisy signal. A smaller ϵ gives a better approximation, however at the cost of less noise reduction. The graphs in figure 5.2 show a noisy function and its convolutions with a smooth function for two values of ϵ; note the tradeoff between smoothness and detail.



(a) A noisy function. (b) A little smoothing. (c) All detail is gone.

Figure 5.2: Using convolution to smooth a noisy function.

The precise sense in which $\varphi_\epsilon * f$ converges to f depends on its regularity and decay. The square-integrable case is the simplest.

Proposition 5.2.2. *Suppose that φ is an absolutely integrable function with*

$$\int_{\mathbb{R}^n} \varphi(\mathbf{x}) d\mathbf{x} = 1.$$

If $f \in L^2(\mathbb{R}^n)$ then $\varphi_\epsilon * f$ converges to f in $L^2(\mathbb{R}^n)$.

Proof. The Plancherel formula implies that

$$\|\varphi_\epsilon * f - f\|_{L^2} = \frac{1}{[2\pi]^{\frac{n}{2}}} \|\widehat{\varphi_\epsilon * f} - \hat{f}\|_{L^2}.$$

The Fourier transform of φ_ϵ is computed using (4.36), it is

$$\mathcal{F}(\varphi_\epsilon)(\boldsymbol{\xi}) = \hat{\varphi}(\epsilon\boldsymbol{\xi}). \tag{5.18}$$

From Theorem 5.1.1 we obtain

$$\|\widehat{\varphi_\epsilon * f} - \hat{f}\|_{L^2} = \|\hat{f}(\hat{\varphi}_\epsilon - 1)\|_{L^2}.$$

The Lebesgue dominated convergence theorem, (5.18) and the fact that $\hat{\varphi}(0) = 1$ lead to the conclusion that

$$\lim_{\epsilon \to \infty} \|\hat{f}(\hat{\varphi}_\epsilon - 1)\|_{L^2} = 0.$$

□

Another important case is the L^1-case

Proposition 5.2.3. *Suppose that φ is an absolutely integrable function with*

$$\int_{\mathbb{R}^n} \varphi(\mathbf{x}) d\mathbf{x} = 1.$$

If $f \in L^1(\mathbb{R}^n)$ then $\varphi_\epsilon * f$ converges to f in the L^1-norm.

Proof. The proof of this result is quite different from the L^2-case it relies on the follwing lemma:

Lemma 5.2.1. *If $f \in L^1(\mathbb{R}^n)$ then*

$$\lim_{\boldsymbol{\tau} \to \mathbf{0}} \|f_{\boldsymbol{\tau}} - f\|_{L^1} = 0.$$

In other words the translation operator, $(\boldsymbol{\tau}, f) \mapsto f_{\boldsymbol{\tau}}$ is a continuous map of $\mathbb{R}^n \times L^1(\mathbb{R}^n)$ to $L^1(\mathbb{R}^n)$. The proof of this statement is left to the exercises. The triangle inequality shows that

$$\begin{aligned}
\|\varphi_\epsilon * f - f\|_{L^1} &= \int\limits_{\mathbb{R}^n} \left| \int\limits_{\mathbb{R}^n} [f(\mathbf{x} - \epsilon\mathbf{t}) - f(\mathbf{x})]\varphi(\mathbf{t})d\mathbf{t} \right| d\mathbf{x} \\
&\leq \int\limits_{\mathbb{R}^n} |\varphi(\mathbf{t})| \left[\int\limits_{\mathbb{R}^n} |f(\mathbf{x} - \epsilon\mathbf{t}) - f(\mathbf{x})|d\mathbf{x} \right] d\mathbf{t} \\
&= \int\limits_{\mathbb{R}^n} |\varphi(\mathbf{t})| \|f_{\epsilon\mathbf{t}} - f\|_{L^1} d\mathbf{t}.
\end{aligned} \tag{5.19}$$

The last integrand is bounded by $2\|f\|_{L^1}|\varphi(\mathbf{t})|$ and therefore the limit, as ϵ goes to zero, can be brought inside the integral. The conclusion of the proposition follows from Lemma. □

It is also useful to examine $\varphi_\epsilon * f(\mathbf{x})$ at points where f is smooth.

Proposition 5.2.4. *Let f be a locally integrable function and suppose that φ has bounded support. If f is continuous at $\mathbf{x}$ then*

$$\lim_{\epsilon \downarrow 0} \varphi_\epsilon * f(\mathbf{x}) = f(\mathbf{x}).$$

Proof. As f is continuous at $\mathbf{x}$, given $\eta > 0$ there is a $\delta > 0$ so that

$$\|\mathbf{x} - \mathbf{y}\| < \delta \Rightarrow |f(\mathbf{x}) - f(\mathbf{y})| < \eta. \tag{5.20}$$

If ϵ is sufficiently small, say less than ϵ_0, then the support of φ_ϵ is contained in the ball of radius δ. Finally since the total integral of φ is 1 we have, for an $\epsilon < \epsilon_0$ that

$$\begin{aligned}
|\varphi_\epsilon * f(\mathbf{x}) - f(\mathbf{x})| &= \left| \int\limits_{B_\delta} \varphi_\epsilon(\mathbf{y})(f(\mathbf{x} - \mathbf{y}) - f(\mathbf{x}))d\mathbf{y} \right| \\
&\leq \int\limits_{B_\delta} \varphi_\epsilon(\mathbf{y})|f(\mathbf{x} - \mathbf{y}) - f(\mathbf{x})|d\mathbf{y} \\
&\leq \int\limits_{B_\delta} \varphi_\epsilon(\mathbf{y})\eta d\mathbf{y} \\
&\leq \eta.
\end{aligned} \tag{5.21}$$

In the second line we use the fact that φ_ϵ is non-negative and, in the third line, estimate (5.20). □

Remark 5.2.2. If f has k continuous derivatives in $B_\delta(\mathbf{x})$ then, for $\epsilon < \delta$ and $\boldsymbol{\alpha}$ with $|\boldsymbol{\alpha}| \leq k$, Proposition 5.2.1 implies that

$$\partial_{\mathbf{x}}^{\alpha}(\varphi_\epsilon * f)(\mathbf{x}) = \varphi_\epsilon * \partial_{\mathbf{x}}^{\boldsymbol{\alpha}} f(\mathbf{x}).$$

Proposition 5.2.4 can be applied to conclude that

$$\lim_{\epsilon \to 0} \partial_{\mathbf{x}}^{\alpha}(\varphi_\epsilon * f)(\mathbf{x}) = \partial_{\mathbf{x}}^{\boldsymbol{\alpha}} f(\mathbf{x}).$$

There are many variants of these results. The main point of the proof is that φ is absolutely integrable. Many similar *looking* results appear in analysis, though with much more complicated proofs. In most of these cases φ is *not* absolutely integrable. For example, the Fourier inversion formula in one-dimension amounts to the statement that $\varphi_\epsilon * f$ converges to f where $\varphi(x) = \pi^{-1}\operatorname{sinc}(x)$. As we have noted several times, before $\operatorname{sinc}(x)$ is not absolutely integrable.

In practice, infinitely differentiable functions can be difficult to work with. To simplify computations a finitely differentiable version may be preferred. For example, given $k \in \mathbb{N}$ define the function

$$\psi_k(x) = \begin{cases} c_k(1-x^2)^k & \text{if } |x| \leq 1, \\ 0 & \text{if } |x| > 1. \end{cases} \tag{5.22}$$

The constant, c_k is selected so that ψ_k has total integral one. The function ψ_k has $k-1$ continuous derivatives. If

$$\psi_{k,\epsilon}(x) = \epsilon^{-1}\psi_{k,\epsilon}(\frac{x}{\epsilon})$$

and f is locally integrable, then $< \psi_{k,\epsilon} * f >$ is a sequence of $k-1$-times differentiable functions, which converge, in an appropriate sense to f.

Using these facts we now complete the proof of the Fourier inversion formula. Thus far Theorems 4.2.1 and 4.5.1 were proved with the additional assumption that f is continuous.

Proof of the Fourier inversion formula, completed. Suppose that f and $\hat{f}$ are absolutely integrable and φ_ϵ is as above. Note that $\hat{f}$ is a continuous function. For each $\epsilon > 0$ the function $\varphi_\epsilon * f$ is absolutely integrable and continuous. Its Fourier transform, $\hat{\varphi}(\epsilon\boldsymbol{\xi})\hat{f}(\boldsymbol{\xi})$, is absolutely integrable. As ϵ goes to zero it converges locally uniformly to $\hat{f}(\xi)$. Since these functions are continuous we can use the Fourier inversion formula to conclude that

$$\varphi_\epsilon * f(\mathbf{x}) = \frac{1}{2\pi}\int\limits_{\mathbb{R}^n} \hat{\varphi}(\epsilon\boldsymbol{\xi})\hat{f}(\boldsymbol{\xi})^{i\langle\mathbf{x},\boldsymbol{\xi}\rangle}d\boldsymbol{\xi}.$$

This is a locally uniformly convergent family of continuous functions and therefore has a continuous limit. The right hand side converges pointwise to

$$F(\mathbf{x}) = \int\limits_{\mathbb{R}^n} \hat{f}(\boldsymbol{\xi})^{i\langle\mathbf{x},\boldsymbol{\xi}\rangle}d\boldsymbol{\xi}.$$

Lemma 5.2.1 implies that $\|\varphi_\epsilon * f - f\|_{L^1}$ also goes to zero as ϵ tends to 0 and therefore $F(x) = f(x)$. (To be precise we should say that after modification on a set of measure 0, $F(x) = f(x)$.) This completes the proof of the Fourier inversion formula. □

Exercises

Exercise 5.2.1. Let f be an integrable function with support in the interval $[a, b]$ and g an integrable function with support in $[-\epsilon, \epsilon]$. Show that the support of $f * g(x)$ is contained in $[a - \epsilon, b + \epsilon]$.

Exercise 5.2.2. Use Corollary A.7.1 to prove Lemma 5.2.1.

Exercise 5.2.3. Give the details of the argument, using Lemma 5.2.1, showing that if f is an L^1-function then

$$\lim_{\epsilon\to 0}\int\limits_{\mathbb{R}^n} \varphi(\mathbf{t})\|f_{\epsilon\mathbf{t}} - f\|_{L^1}d\mathbf{t} = 0.$$

Exercise 5.2.4. Use the method used to prove Proposition 5.2.4 to show that if $f \in L^p(\mathbb{R})$ for a $1 \leq p < \infty$ then $\varphi_\epsilon * f$ converges to f in the L^p-norm. Give an example to show that if f is a bounded, though discontinuous function, then $\|\varphi_\epsilon * f - f\|_\infty$ may fail to tend to zero.

Exercise 5.2.5. Let $\psi_\epsilon(x) = [2\epsilon]^{-1}\chi_{[-\epsilon,\epsilon]}(x)$. Show that if $f \in L^2(\mathbb{R})$ then $\psi_\epsilon * f$ converges to f in $L^2(\mathbb{R})$.

Exercise 5.2.6. For the functions ψ_k, defined in (5.22), find the constants c_k so that

$$\int_{-1}^{1} \psi_k(x)dx = 1.$$

Exercise 5.2.7. Use the Fourier inversion formula to prove that

$$\widehat{fg}(\xi) = \frac{1}{2\pi}\hat{f} * \hat{g}(\xi). \tag{5.23}$$

What assumptions are needed for $\hat{f} * \hat{g}$ to make sense?

Exercise 5.2.8. * For k a positive integer suppose that f and $\xi^k \hat{f}(\xi)$ belong to $L^2(\mathbb{R})$. By approximating f by smooth functions of the form $\varphi_\epsilon * f$ show that f has k L^2-derivatives.

5.2.2 The support of $f * g$.

Suppose that f and g have bounded support. For applications to medical imaging it is important to understand how the support of $f * g$ is related to the supports of f and g. To that end we define the *algebraic sum* of two subsets of $\mathbb{R}^n$.

Definition 5.2.1. Suppose A and B are subsets of $\mathbb{R}^n$. The algebraic sum of A and B is defined as the set

$$A + B = \{\mathbf{a} + \mathbf{b} \in \mathbb{R}^n \ : \ \mathbf{a} \in A, \text{ and } \mathbf{b} \in B\}.$$

Using this concept we can give a quantitative result describing the way in which convolution “smears” out the support of a function.

Lemma 5.2.2. *The support of $f * g$ is contained in* $\operatorname{supp} f + \operatorname{supp} g$.

Proof. Suppose that $\mathbf{x}$ is not in $\operatorname{supp} f + \operatorname{supp} g$. This means that no matter which $\mathbf{y}$ is selected either $f(\mathbf{y})$ or $g(\mathbf{x} - \mathbf{y})$ is zero. Otherwise $\mathbf{x} = \mathbf{y} + (\mathbf{x} - \mathbf{y})$ would belong to $\operatorname{supp} f + \operatorname{supp} g$. This implies that $f(\mathbf{y})g(\mathbf{x} - \mathbf{y})$ is zero for all $\mathbf{y} \in \mathbb{R}^n$ and therefore

$$f * g(\mathbf{x}) = \int_{\mathbb{R}^n} f(\mathbf{y})g(\mathbf{x} - \mathbf{y})d\mathbf{y} = 0$$

as well. This proves the lemma. □

Suppose that f is a function which represents an image. For example we could imagine that f takes values between 0 and 1 with 0 corresponding to white and 1 to black. Values in between correspond to shades of grey. Convolution with a function like φ_ϵ, defined in section 5.2.1, provides a reasonable model for the measurement of such an image. Suppose that f has bounded support then, by Lemma 5.2.2,

$$\operatorname{supp}(f * \varphi_\epsilon) \subset \operatorname{supp} f + \operatorname{supp} \varphi_\epsilon.$$

As φ_ϵ is supported in B_ϵ this is the ϵ-neighborhood of the support of f,

$$\operatorname{supp}(f)_\epsilon = \{\mathbf{x} \in \mathbb{R}^n \,:\, \operatorname{dist}(\mathbf{x}, \operatorname{supp}(f)) \leq \epsilon\}.$$

The figures indicates what happens in the 2-dimensional case.

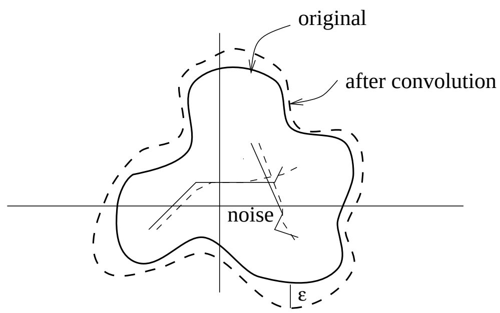


Figure 5.3: f is smeared into the ϵ-neighborhood of $\operatorname{supp}(f)$.

The convolution of f with φ_ϵ smears the image represented by f. The value of $\varphi_\epsilon * f(\mathbf{x})$ depends on the values of f in the ball of radius ϵ about $\mathbf{x}$, hence parameter ϵ reflects the resolution of the measuring apparatus.

At points where the image is slowly varying the measured image is close to the actual image. Near points where f is rapidly varying this may not be the case. Noise is usually a high frequency phenomenon with "mean zero;" smoothing averages out the noise. At the same time, the image is blurred. The size of ϵ determines the degree of blurring. The Fourier transform of $f * \varphi_\epsilon$ is $\hat{f}(\boldsymbol{\xi})\hat{\varphi}(\epsilon\boldsymbol{\xi})$. Because φ has integral 1 it follows that $\hat{\varphi}(0) = 1$. As the support of φ is a bounded set, its Fourier transform is a smooth function. This shows that $\hat{\varphi}(\epsilon\boldsymbol{\xi}) \approx 1$ if $\|\boldsymbol{\xi}\| << \epsilon^{-1}$. Thus the "low frequency" part of $f * \varphi_\epsilon$ closely approximates the low frequency part of f. On the other hand $\hat{\varphi}(\boldsymbol{\xi})$ tend to zero rapidly as $\|\boldsymbol{\xi}\| \to \infty$ and therefore the high frequency content of f is suppressed in $f * \varphi_\epsilon$. Unfortunately both noise *and* fine detail are carried by the high frequency components. Using convolution to supress noise inevitably destroys fine detail.

5.2.3 Convolution equations

Convolution provides a model for many measurement and filtering processes. If f is the state of a system then, for a fixed function ψ, the output g is modeled by the convolution $g = \psi * f$. In order to recover the state of the system from the output one must therefore *solve* this equation for f as a function of g. Formally this equation is easy to solve, (5.10) implies that

$$\hat{f}(\boldsymbol{\xi}) = \frac{\hat{g}(\boldsymbol{\xi})}{\hat{\psi}(\boldsymbol{\xi})}.$$

There are several problems with this approach. The most obvious problem is that $\hat{\psi}$ may vanish for some values of $\boldsymbol{\xi}$. If the model were perfect then, of course, $\hat{g}(\boldsymbol{\xi})$ would also have to vanish at the same points. In real applications this leads to serious problems with stability. A second problem is that, if $\psi(\mathbf{x})$ is absolutely integrable, then the Riemann-Lebesgue lemma implies that $\hat{\psi}(\boldsymbol{\xi}) \to 0$ as $\|\boldsymbol{\xi}\| \to \infty$. Unless the measurement g is very smooth and noise free we would be unable to invert the Fourier transform of this ratio to determine f. In Chapter 9 we discuss how these issues are handled in practice.

Example 5.2.1. The rectangle function defines a simple weight, $\psi_\epsilon = (2\epsilon)^{-1}\chi_{[-\epsilon,\epsilon]}$. Its Fourier transform is given by

$$\hat{\psi}_\epsilon(\xi) = \operatorname{sinc}(\epsilon\xi).$$

This function has zeros at $\xi = \pm(\epsilon^{-1}m\pi)$, where m is any positive integer. These zeros are isolated so it seems reasonable that an integrable function f should be uniquely specified by the averages $\psi_\epsilon * f$, for any $\epsilon > 0$. In fact it is, but the problem of reconstucting f from these averages is not well posed.

Example 5.2.2. Suppose that ψ is a non-negative function which vanishes outside the interval $[-\epsilon,\epsilon]$ and has total integral 1,

$$\int\limits_{-\infty}^{\infty} \psi(x)dx = 1.$$

If f is a locally integrable function then $f * \psi(x)$ is the weighted average of the values of f over the interval $[x-\epsilon, x+\epsilon]$. Note that $\psi * \psi$ also has total integral 1

$$\begin{aligned}\int\limits_{-\infty}^{\infty} \psi * \psi(x)dx &= \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} \psi(y)\psi(x-y)dydx \\ &= \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} \psi(y)\psi(t)dtdy \\ &= 1\cdot 1 = 1.\end{aligned} \tag{5.24}$$

In the second to last line we reversed the order of the integrations and set $t = x - y$.

Thus $f * (\psi * \psi)$ is again an average of f. Note that $\psi * \psi(x)$ is generally non-zero for $x \in [-2\epsilon, 2\epsilon]$, so convolving with $\psi * \psi$ produces more blurring than convolution with ψ alone. Indeed we know from the associativity of the convolution product that

$$f * (\psi * \psi) = (f * \psi) * \psi,$$

so we are averaging the averages, $f * \psi$. This can be repeated as many times as one likes, the j-fold convolution $\psi *_j \psi$ has total integral 1 and vanishes outside the interval $[-j\epsilon, j\epsilon]$. Of course the Fourier transform of $\psi *_j \psi$ is $[\hat{\psi}(\xi)]^j$ which therefore decays j times as fast as $\hat{\psi}(\xi)$.

We could also use the *scaled* j-fold convolution $\delta^{-1}\psi *_j \psi(\delta^{-1}x)$ to average our data. This function vanishes outside the interval $[-j\delta\epsilon, j\delta\epsilon]$ and has Fourier transform $[\hat{\psi}(\delta\xi)]^j$. If we choose $\delta = j^{-1}$ then convolving with this function will not blur details any more than convolving with ψ itself but better suppresses high frequency noise. By choosing j and δ we can control, to some extent, the trade off between blurring and noise suppression.

Exercises

Exercise 5.2.9. If a and b are positive numbers then define

$$w_{a,b}(x) = \frac{1}{2}[r_a(x) + r_b(x)].$$

Graph $w_{a,b}(x)$ for several different choices of (a, b). Show that for appropriate choices of a and b the Fourier transform $\hat{w}_{a,b}(\xi)$ does not vanish for any value of ξ.

Exercise 5.2.10. Define a function

$$f(x) = \chi_{[-1,1]}(x)(1 - |x|)^2.$$

Compute the Fourier transform of this function and show that it does not vanish anywhere. Let $f_j = f *_j f$ (the j-fold convolution of f with itself). Show that the Fourier transforms, $\hat{f}_j$ are also non-vanishing.

5.3 The δ-function

See: A.5.6.

The convolution product defines a multiplication on $L^1(\mathbb{R}^n)$ with all the usual properties of a product except that there is no unit. If i were a unit then $i * f = f$ for every function in $L^1(\mathbb{R}^n)$. Taking the Fourier transform, this would imply that, for every $\boldsymbol{\xi}$,

$$\hat{i}(\boldsymbol{\xi})\hat{f}(\boldsymbol{\xi}) = \hat{f}(\boldsymbol{\xi}).$$

This shows that $\hat{i}(\boldsymbol{\xi}) \equiv 1$ and therefore i cannot be an L^1-function. Having a multiplicative unit is so useful that engineers, physicists and mathematicians have all found it necessary to simply define one. It is called the *δ-function* and is defined by the property that for any continuous function f

$$f(0) = \int_{\mathbb{R}^n} \delta(\mathbf{y}) f(\mathbf{y}) d\mathbf{y}. \tag{5.25}$$

Proceeding formally we see that

$$\begin{aligned} \delta * f(\mathbf{x}) &= \int_{\mathbb{R}^n} \delta(\mathbf{y}) f(\mathbf{x} - \mathbf{y}) d\mathbf{y} \\ &= f(\mathbf{x} - \mathbf{0}) = f(\mathbf{x}). \end{aligned} \tag{5.26}$$

So at least for continuous functions $\delta * f = f$.

It is important to remember *the δ-function is not a function.* In the mathematics literature the δ-function is an example of a *distribution* or *generalized function.* The basic properties of generalized functions are introduced in Appendix A.5.6. In the engineering and physics literature it is sometimes called a *unit impulse.* In section 4.4.4 the Fourier transform is extended to generalized functions (at least in the one-dimensional case). The Fourier transform of δ is as expected, identically equal to 1:

$$\mathcal{F}(\delta) \equiv 1.$$

While (5.25) only makes sense for functions continuous at 0, the convolution of δ with arbitrary locally integrable functions is well defined and satisfies $\delta * f = f$. This is not too different from the observation that if f and g are L^1-functions then $f * g(\mathbf{x})$ may not be defined at every point, nonetheless, $f * g$ is a well defined element of $L^1(\mathbb{R}^n)$.

In both mathematics and engineering it is useful to have approximations for the δ-function. There are two complementary approaches to this problem, one is to use functions like φ_ϵ, defined in (5.14) to approximate δ in $\mathbf{x}$-space. The other is to approximate $\hat{\delta}$ in $\boldsymbol{\xi}$-space. To close this chapter we formalize the concept of resolution and considering some practical aspects of approximate δ-functions in one-dimension.

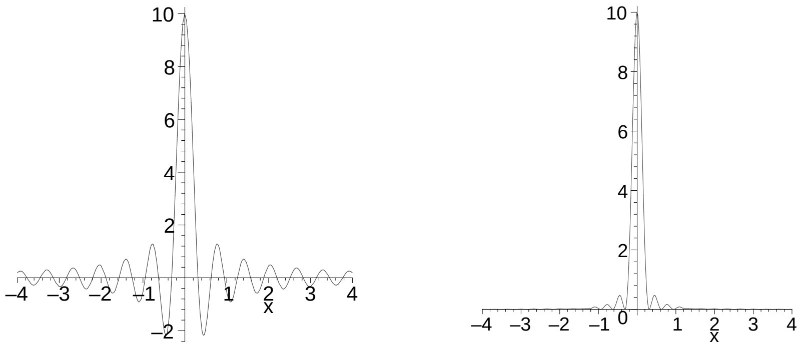
5.3.1 Approximating the δ-function in 1-dimension

Suppose that φ is an even function with bounded support. The Fourier transform of φ_ϵ is $\hat{\varphi}(\epsilon\xi)$. Because φ_ϵ vanishes outside a finite interval its Fourier transform is a smooth function and $\hat{\varphi}(0) = 1$. As φ is a non-negative, even function its Fourier transform is real valued and assumes its maximum at zero. In applications it is important that the difference $1 - \hat{\varphi}(\epsilon\xi)$ remain small over a specified interval $[-B, B]$. It is also important that $\hat{\varphi}(\epsilon\xi)$ tend to zero rapidly outside a somewhat larger interval. As φ is non-negative, $\partial_\xi\hat{\varphi}(0) = 0$; this means that the behavior of $\hat{\varphi}(\xi)$ for ξ near to zero is largely governed by the "second moment"

$$\partial_\xi^2\hat{\varphi}(0) = -\int_{-\infty}^{\infty} x^2\varphi(x)dx.$$

One would like this number to be small. This is accomplished by putting more of the mass of φ near to $x = 0$. On the other hand the rate at which $\hat{\varphi}$ decays as $|\xi| \to \infty$ is determined by the smoothness of φ. If $\varphi = \frac{1}{2}\chi_{[-1,1]}$ then $\hat{\varphi}$ decays like $|\xi|^{-1}$. Better decay is obtained by using a smoother function. In applications having $\hat{\varphi}$ absolutely integrable is usually adequate. In one-dimension this is the case if φ is continuous and piecewise differentiable.

The other approach to constructing approximations to the δ-function is to approximate its Fourier transform. One uses a sequence of functions which are approximately 1 in an interval $[-B, B]$ and vanish outside a larger interval. Again a simple choice is $\chi_{[-B,B]}(\xi)$. The inverse Fourier transform of this function is $\psi_B(x) = \pi^{-1}B\operatorname{sinc}(Bx)$. In this context it is called a *sinc pulse*. Note that ψ_B assumes both positive and negative values. A sinc-pulse is not absolutely integrable, the fact that the improper integral of ψ_B over the whole real line equals 1 relies on subtle cancellations between the positive and negative parts of the integral. Because ψ_B is not absolutely integrable, it is often a poor choice for approximating the δ-function. Approximating $\hat{\delta}$ by $(2B)^{-1}\chi_{[-B,B]} * \chi_{[-B,B]}(\xi)$ gives a sinc^2-pulse, $(2B)^{-1}\psi_B^2(x)$, as an approximation to δ. This function has better properties: it does not assume negative values, is more sharply peaked at 0 and is absolutely integrable. These functions are graphed in figure 5.4.

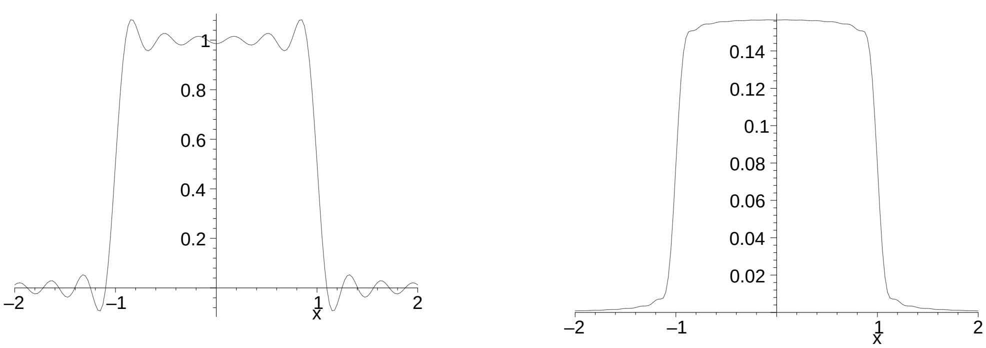


(a) A sinc-pulse.

(b) A sinc^2-pulse.

Figure 5.4: Approximate δ-functions

Neither the sinc nor sinc^2 has bounded support, both functions have oscillatory "tails" extending to infinity. In the engineering literature these are called *side lobes.* Side lobes result from the fact that the Fourier transform vanishes outside a bounded interval, see section 4.4.3. The convolutions of these functions with $\chi_{[-1,1]}$ are shown in figure 5.5. In figure 5.5(a) notice that the side lobes produce large oscillations near the jump. This is an example of the "Gibbs phenomenon." It results from using a discontinuous cutoff function in the Fourier domain. This effect is analyzed in detail, for the case of Fourier series in section 7.5.



(a) $\text{sinc} * \chi_{[-1,1]}$.

(b) $\text{sinc}^2 * \chi_{[-1,1]}$.

Figure 5.5: Approximate δ-functions convolved with $\chi_{[-1,1]}$.

Exercises

Exercise 5.3.1. Suppose that f is a continuous L^1-function and φ is absolutely integrable with $\int_{\mathbb{R}} \varphi = 1$. Show that $< \varphi_\epsilon * f >$ converges pointwise to f.

Exercise 5.3.2. Suppose that φ is an integrable function on the real line with total integral 1 and f is an integrable function such that, for a $k > 1$,

$$|\hat{f}(\xi)| \leq \frac{C}{(1 + |\xi|)^k}.$$

Use the Fourier inversion formula to estimate the error $\|\varphi_\epsilon * f(x) - f(x)|$.

5.3.2 Resolution and the full width half maximum

We now give a standard definition for the resolution present in a measurement of the form $\psi * f$. Resolution is a subtle and, in some senses, subjective concept. It is mostly useful for purposes of comparision. The definition presented here is just one of many possible definitions.

Suppose that ψ is a non-negative function with a single hump similar to those shown in figure 5.1. The important features of this function are

1. It is non-negative,
2. It has a single maximum value, which it attains at 0,
3. It is monotone increasing to the left of the maximum and monotone decreasing to the right.

(5.27)

Definition 5.3.1. Let ψ satisfy these conditions and let M be the maximum value it attains. Let $x_1 < 0 < x_2$ be respectively the smallest and largest numbers so that

$$\psi(x_1) = \psi(x_2) = \frac{M}{2}.$$

The difference $x_2 - x_1$ is called the *full width half maximum* of the function ψ. It is denoted FWHM(ψ). If f is an input then the resolution available in the measurement, $\psi * f$ is *defined* to be the FWHM(ψ).

In principle if FWHM(ψ_1) < FWHM(ψ_2) then $f \mapsto \psi_1 * f$ should have better resolution than $f \mapsto \psi_2 * f$. Here is a heuristic explanation for this definition. Suppose that the signal f is pair of unit impulses separated by a distance d,

$$f(x) = \delta(x) + \delta(x - d).$$

Convolving ψ with f produces two copies of ψ,

$$\psi * f(x) = \psi(x) + \psi(x - d).$$

If $d >$ FWHM(ψ) then $\psi * f$ has two distinct maxima separated by a valley. If $d \leq$ FWHM(ψ) then the distinct maxima disappear. If the distance between the impulses is greater than the FWHM(ψ) then we can "resolve" them in the filtered output. More generally the FWHM(ψ) is considered to be the smallest distance between distinct "features" in f which can be seen in $\psi * f$. In figure 5.6 we use a triangle function for ψ. The FWHM of this function is 1, the graphs show ψ and the results of convolving ψ with a pair of unit impulses separated, respectively by $1.2 > 1$ and $.8 < 1$.

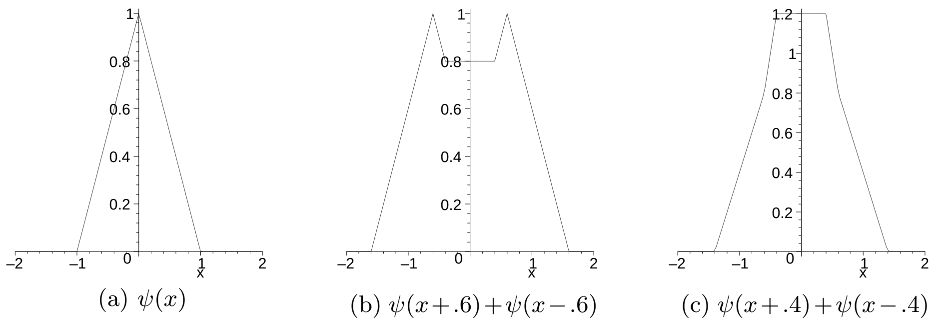


Figure 5.6: Illustration of the FWHM definition of resolution

This definition is often extended to functions which do not satisfy all the conditions in (5.27) but are qualitatively similar. For example the characteristic function of an interval $\chi_{[-B,B]}(x)$ has a unique maximum value and is monotone to the right and left of the maximum. The FWHM($\chi_{[-B,B]}$) is therefore $2B$. Another important example is the sinc-function. It has a unique maximum and looks correct near to it. This function also has large side-lobes which considerably complicate the behavior of the map $f \mapsto f * \text{sinc}$. The FWHM(sinc) is taken to be the full width half maximum of its central peak, it is approximately given by

$$\text{FWHM}(\text{sinc}) \approx 1.895494.$$

We return to the problem of quantifying resolution in Chapter 9.

Exercises

Exercise 5.3.3. Numerically compute the FWHM($\text{sinc}^2(x)$). How does it compare to FWHM($\text{sinc}(x)$).

Exercise 5.3.4. Suppose that

$$h_j(x) = \left[\frac{\sin(x)}{x}\right]^j.$$

Using the Taylor expansion for sine function show that, as j gets large,

$$\text{FWHM}(h_j) \simeq \sqrt{\frac{6 \log 2}{j}}.$$

Exercise 5.3.5. Using the Taylor expansion for the sine, show that as B gets large

$$\text{FWHM}(\text{sinc}(Bx)) \simeq \frac{\sqrt{3}}{B}.$$

Exercise 5.3.6. For $a > 0$ let $g_a(x) = e^{-\frac{x^2}{a^2}}$. Compute FWHM($g_a$) and FWHM($g_a * g_b$).