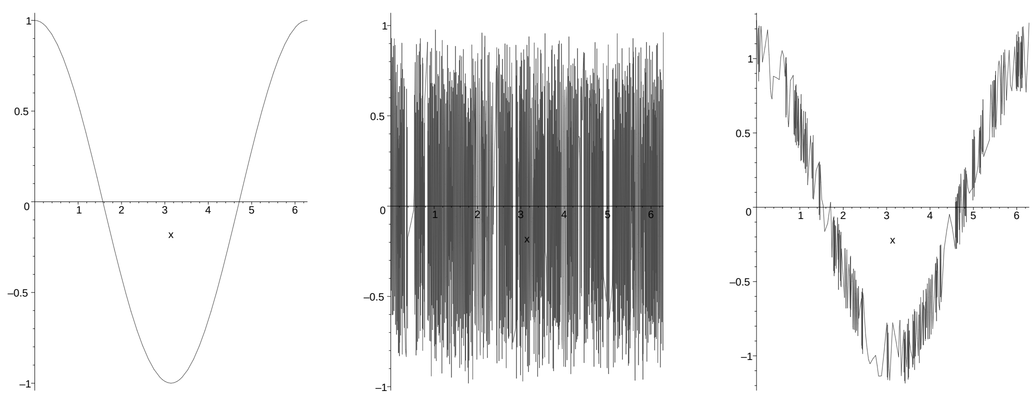
Chapter 5

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.* The idea for convolution comes from considering *moving averages.*

Suppose we would like to analyze a smooth function of one variable, s but the available data is contaminated by noise. For the purposes of the present discussion, this means that the measured signal is of the form $f = s + \epsilon n$. Here ϵ is a (small) number and n is function which models the noise. An example is shown in figure 5.1.



(a) A smooth function s.

(b) A typical noise function n.

(c) A realistic measurement $s + .25n$.

Figure 5.1: A smooth function gets corrupted by noise.

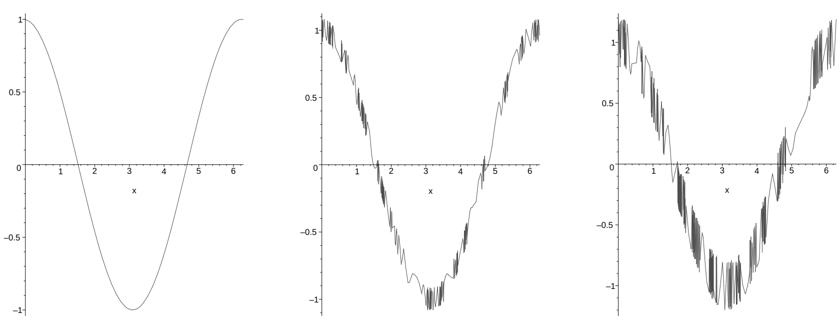
Noise is typically represented by a rapidly varying function which is locally of "mean zero." This means that, for any x, and a large enough δ, the average

$$\frac{1}{\delta} \int\limits_{x}^{x+\delta} n(y)dy$$

is small compared to the size of n. The more "random" the noise, the smaller δ can be taken. On the other hand, since s is a smooth function, the analogous average of s should be close to $s(x)$. The moving average of f is defined to be

$$\mathcal{M}_\delta(f)(x) = \frac{1}{\delta}\int_x^{x+\delta} f(y)dy. \tag{5.1}$$

If the noise is very random, so that δ can be taken small, then $\mathcal{M}_\delta(f)$ should be close to s. The results of applying this averaging process to the function shown in figure 5.1(c) are shown in figure 5.2.



(a) $\delta = .1$. (b) $\delta = .001$ (c) $\delta = .00001$

Figure 5.2: The moving average $\mathcal{M}_\delta(f)$ for various values of δ.

There is a somewhat neater and more flexible way to express the operation defined in (5.1). Define the weight function

$$m_\delta(x) = \begin{cases} \frac{1}{\delta} & \text{for } x \in [-\delta, 0], \\ 0 & \text{otherwise.} \end{cases}$$

The moving average then becomes

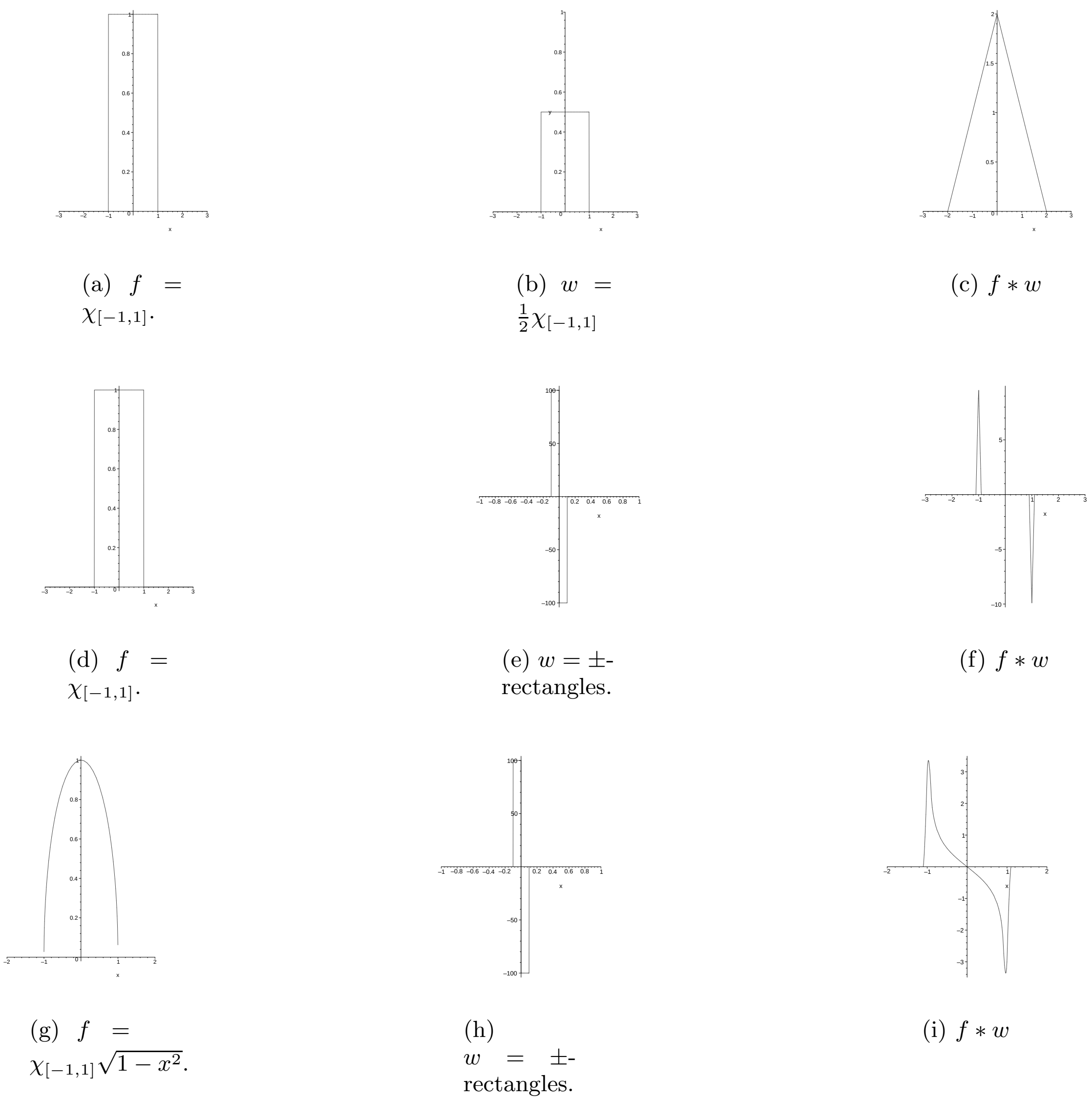
$$\mathcal{M}_\delta(f) = \int_{-\infty}^{\infty} f(y)m_\delta(x-y)dy. \tag{5.2}$$

In this formulation we see that the value of $\mathcal{M}_\delta(f)(x)$ is obtained by translating the weight function along the axis, multiplying it by f and integrating. To be a little more precise the weight function is first reflected around the vertical axis, i.e. $m_\delta(y)$ is replaced by $m_\delta(-y)$ and then translated to give $m_\delta(-(y-x)) = m_\delta(x-y)$. At this stage it is a little difficult to motivate the reflection step, but in the end it leads to a much simpler theory.

The weight function in (5.2) is just one possible choice. Depending upon the properties of the noise (or the signal) it might be advantageous to use a different weight function. For w an integrable function, define the w-weighted moving average by

$$\mathcal{M}_w(f) = \int_{-\infty}^{\infty} f(y)w(x-y)dy. \tag{5.3}$$

For this to be an "average" in the usual sense, w should be non-negative with total integral equal to one. The operation $f \mapsto \mathcal{M}_w(f)$ is defined under much weaker conditions on w. The main features of this operation are: 1. It is linear in f. 2. The weight assigned to $f(y)$ in the ouput $\mathcal{M}_w(f)(x)$ depends only on the difference $x - y$. Many operations of this type appear in mathematics and image processing.



(a) $f = \chi_{[-1,1]}$.

(b) $w = \frac{1}{2}\chi_{[-1,1]}$

(c) $f * w$

(d) $f = \chi_{[-1,1]}$.

(e) $w = \pm$-rectangles.

(f) $f * w$

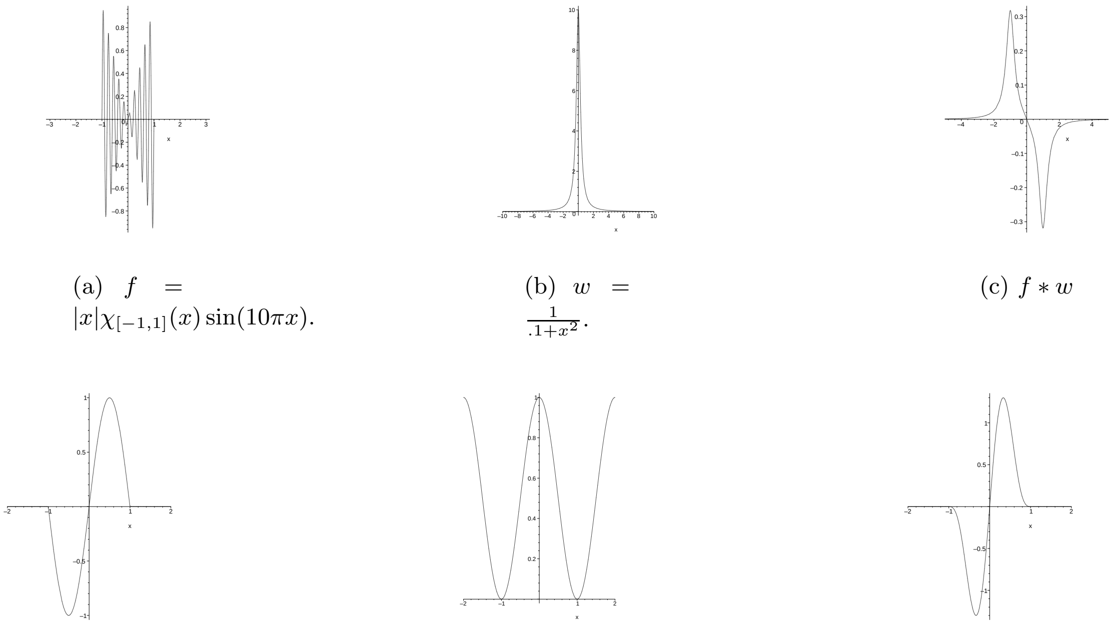
(g) $f = \chi_{[-1,1]}\sqrt{1-x^2}$.

(h) $w = \pm$-rectangles.

(i) $f * w$

Figure 5.3: Examples of the convolution product.

It turns out that the simplest theory results from thinking of this as a bi-linear operation in the two functions f and w. The result, denoted by $f * w$, is called the *convolution product*. Several examples are shown in figures 5.3 and 5.4. As we show below, this operation has many of the properties of ordinary pointwise multiplication, with one important addition: Convolution is intimately connected to the Fourier transform. Because there are very efficient algorithms for approximating the Fourier transform and its inverse, convolution lies



(a) $f = |x|\chi_{[-1,1]}(x)\sin(10\pi x)$.

(b) $w = \frac{1}{.1+x^2}$.

(c) $f * w$

(d) $f = \chi_{[-1,1]}(x)\sin(\pi x)$.

(e) $w = \cos^2(\frac{\pi x}{2})$

(f) $f * w$

Figure 5.4: More examples of the convolution product.

at the heart of many practical filters. After defining the convolution product for functions 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. As the definition and formal properties of this operation do not depend on the dimension, we 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 a bounded, locally integrable function then the *convolution product* of f and g is the function on $\mathbb{R}^n$ defined by the integral

$$f * g(\mathbf{x}) = \int\limits_{\mathbb{R}^n} f(\mathbf{x}-\mathbf{y})g(\mathbf{y})d\mathbf{y}. \tag{5.4}$$

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 $L^2(\mathbb{R}^n)$.

We consider a couple of additional examples:

Example 5.1.1. 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. For any locally integrable function f the value of $f * g(\mathbf{x})$ is given by

$$\begin{aligned} f * g(\mathbf{x}) &= \int\limits_{\mathbb{R}^n} f(\mathbf{x}-\mathbf{y})g(\mathbf{y})d\mathbf{y} \\ &= \frac{c_n}{r^n}\int\limits_{B_r} f(\mathbf{x}-\mathbf{y})d\mathbf{y}. \end{aligned} \tag{5.5}$$

This is the ordinary 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 5.1.2. 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

$$S_R(f)(x) = \frac{1}{2\pi}\int\limits_{-R}^{R} \hat{f}(\xi)e^{ix\xi}d\xi. \tag{5.6}$$

This can be represented as a convolution,

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

where

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

For functions in either L^1 or L^2 this convolution is given by an absolutely convergent integral.

5.1.1 Basic properties of the convolution product

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

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

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*

$$\|f * g\|_{L^1} \leq \|f\|_{L^1}\|g\|_{L^1}. \tag{5.8}$$

Proof. It follows from the triangle inequality that

$$\begin{aligned}\int_{\mathbb{R}^n} |f * g(\mathbf{x})| d\mathbf{x} &\leq \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} |f(\mathbf{x}-\mathbf{y})g(\mathbf{y})| d\mathbf{y} d\mathbf{x} \\ &= \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} |f(\mathbf{x}-\mathbf{y})g(\mathbf{y})| d\mathbf{x} d\mathbf{y}.\end{aligned} \tag{5.9}$$

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 $\mathbb{R}^n \times \mathbb{R}^n$. Changing variables in the $\mathbf{x}$-integral by setting $\mathbf{t} = \mathbf{x} - \mathbf{y}$, we get

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

□

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 (5.8). 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* (5.8).

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

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

might diverge, but only for $\mathbf{x}$ belonging to a set of measure zero. An inequality analogous to (5.8) 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

$$\|f * g\|_{L^p} \leq \|f\|_{L^p}\|g\|_{L^1}. \tag{5.10}$$

The proof of this statement is left to the exercises.

Example 5.1.3. Some decay conditions are required for $f * g$ to be defined. If $f(x) = [\sqrt{1+|x|}]^{-1}$ then

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

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. *The convolution product is commutative, distributive and associative, that is if f_1, f_2, f_3 belong to $L^1(\mathbb{R}^n)$ then the following identities hold:*

$$\begin{aligned} f_1 * f_2 &= f_2 * f_1, \\ f_1 * (f_2 + f_3) &= f_1 * f_2 + f_1 * f_3, \\ f_1 * (f_2 * f_3) &= (f_1 * f_2) * f_3. \end{aligned} \tag{5.11}$$

Remark 5.1.3. If convolution were defined without "reflecting" the argument of the second function through the origin, i.e. if $f * g(\mathbf{x}) = \int f(\mathbf{y})g(\mathbf{y}-\mathbf{x})d\mathbf{y}$ then the convolution product would not be commutative. Instead we would have the identity $f * g(\mathbf{x}) = g * f(-\mathbf{x})$.

Proof. We prove the first assertion; it suffices to assume that f_2 is bounded, the general case then follows from (5.8). The definition states that

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

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

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

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

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 see that such a function cannot exist. For if

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

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 L^1-functions then*

$$\mathcal{F}(f * g) = \mathcal{F}(f)\mathcal{F}(g). \tag{5.12}$$

Proof. The convolution, $f * g$ is an L^1-function and therefore has a Fourier transform. Because $f(\mathbf{x} - \mathbf{y})g(\mathbf{y})$ is an L^1-function of $(\mathbf{x}, \mathbf{y})$, the following manipulations are easily justified,

$$\begin{aligned} \mathcal{F}(f * g)(\boldsymbol{\xi}) &= \int_{\mathbb{R}^n} (f * g)(\mathbf{x}) e^{-i\langle \boldsymbol{\xi}, \mathbf{x} \rangle} d\mathbf{x} \\ &= \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} f(\mathbf{x} - \mathbf{y}) g(\mathbf{y}) e^{-i\langle \boldsymbol{\xi}, \mathbf{x} \rangle} d\mathbf{y} d\mathbf{x} \\ &= \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} f(\mathbf{t}) g(\mathbf{y}) e^{-i\langle \boldsymbol{\xi}, (\mathbf{y} + \mathbf{t}) \rangle} d\mathbf{t} d\mathbf{y} \\ &= \hat{f}(\boldsymbol{\xi}) \hat{g}(\boldsymbol{\xi}). \end{aligned} \tag{5.13}$$

□

Remark 5.1.4. 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 5.1.4. Let $f = \chi_{[-1,1]}$. Formula (5.12) simplifies the computation of the Fourier transform for $f * f$ or even the j-fold convolution of f with itself

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

In this case

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

Example 5.1.5. A partial inverse for the Fourier transform in n-dimensions is defined by

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

The Fourier transform of the function

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

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, (5.8) 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. Suppose that the convolution product were defined by $f*g(\mathbf{x}) = \int f(\mathbf{y})g(\mathbf{y}-\mathbf{x})d\mathbf{x}$ show that (5.12) would not hold. What would replace it?

Exercise 5.1.9. 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 5.1.10. 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.

Exercise 5.1.11. Use the Fourier inversion formula to prove that

$$\widehat{fg}(\xi) = \frac{1}{2\pi} \hat{f} * \hat{g}(\xi). \tag{5.14}$$

What assumptions are needed for $\hat{f} * \hat{g}$ to make sense?

5.1.2 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 (5.7) shows that any filter defined by convolution with an L^1-function is stable. Indeed the estimates in (5.10) show 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.

Example 5.1.6. Suppose the $\boldsymbol{\tau}$ is a point in $\mathbb{R}^n$; the shift operation $f \mapsto f_{\boldsymbol{\tau}}$ defines a shift invariant filter.

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

Proof. The proof is a simple change of variables.

$$\begin{aligned} C_\psi(f_{\boldsymbol{\tau}})(\mathbf{x}) &= \int\limits_{\mathbb{R}^n} \psi(\mathbf{x}-\mathbf{y})f(\mathbf{y}-\boldsymbol{\tau})d\mathbf{y} \\ &= \int\limits_{\mathbb{R}^n} \psi(\mathbf{x}-\boldsymbol{\tau}-\mathbf{w})f(\mathbf{w})d\mathbf{w} \\ &= C_\psi(f)(\mathbf{x}-\boldsymbol{\tau}). \end{aligned} \tag{5.15}$$

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

$$C_\psi(f) = \mathcal{F}^{-1}(\hat{\psi}\hat{f}). \tag{5.16}$$

This is significant because, as noted above, 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 5.1.12. 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: $\mathcal{T}_r(f)(x) \stackrel{d}{=} f(-x)$.

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

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

Exercise 5.1.13. Suppose that $\mathcal{A}$ and $\mathcal{B}$ are shift invariant. Show that their composition $\mathcal{A} \circ \mathcal{B}(f) \stackrel{d}{=} \mathcal{A}(\mathcal{B}(f))$ is also shift invariant.

5.1.3 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, (5.13) 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 ψ is absolutely integrable, then the Riemann-Lebesgue lemma implies that $\hat{\psi}(\boldsymbol{\xi})$ tends to 0 as $\|\boldsymbol{\xi}\|$ goes to infinity Unless the measurement g is smooth and noise free it is not possible to exactly determine f by applying the inverse Fourier transform to this ratio. In Chapter 9 we discuss how these issues are handled in practice.

Example 5.1.8. The rectangle function defines a simple weight, $\psi_\epsilon = (2\epsilon)^{-1}\chi_{[-\epsilon,\epsilon]}$. Its Fourier transform is given by

$$\hat{\psi}_\epsilon(\xi) = \text{sinc}(\epsilon\xi).$$

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 f cannot be *stably* reconstructed from these averages.

Exercises

Exercise 5.1.14. If a and b are positive numbers then define

$$w_{a,b}(x) = \frac{1}{2}\left[\frac{\chi_{[-a,a]}(x)}{2a} + \frac{\chi_{[-b,b]}(x)}{2b}\right].$$

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 5.1.15. 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.

5.2 Convolution and regularity

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

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

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

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

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

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

$$\partial_{\mathbf{x}}^{\boldsymbol{\alpha}}(f * \varphi) = f * (\partial_{\mathbf{x}}^{\boldsymbol{\alpha}}\varphi). \tag{5.17}$$

Remark 5.2.1. This result is also reasonable from the point of view of the Fourier transform. Suppose that φ has k integrable derivatives, then Proposition 4.5.3 shows that

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

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

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

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

5.2.1 Approximation by smooth functions

Convolution provides a general method for approximating integrable (or locally integrable) functions by smooth functions. Beyond that it gives a technique to define regularized derivatives for functions which are not differentiable. We begin with a definition:

Definition 5.2.1. For a function φ, defined on $\mathbb{R}^n$, and ϵ, a positive real number, define the *scaled function* φ_ϵ by

$$\varphi_\epsilon(\mathbf{x}) = \epsilon^{-n}\varphi(\frac{\mathbf{x}}{\epsilon}). \tag{5.18}$$

While this notation is quite similar to that used, in definition 5.1.2, for the translation of a function, the meaning should be clear from the context. A one-dimensional example is shown in figure 5.5.

Let φ be an infinitely differentiable function with total integral one:

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

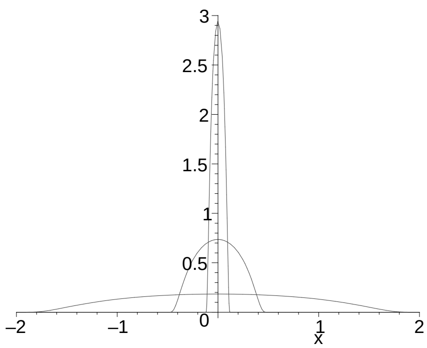


Figure 5.5: Graphs of φ_ϵ, with $\epsilon = .5, 2, 8$.

If φ is supported in the ball of radius 1 then φ_ϵ is supported in the ball of radius ϵ and also has total integral one: Using the change of variables $\epsilon\mathbf{y} = \mathbf{x}$ gives

$$\int\limits_{\mathbb{R}^n} \varphi_\epsilon(\mathbf{x})d\mathbf{x} = \int\limits_{\mathbb{R}^n} \frac{1}{\epsilon^n}\varphi(\frac{\mathbf{x}}{\epsilon})d\mathbf{x} = \int\limits_{\mathbb{R}^n} \varphi(\mathbf{y})d\mathbf{y} = 1. \tag{5.19}$$

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

$$\varphi_\epsilon * f(\mathbf{x}) - f(\mathbf{x}) = \int\limits_{B_\epsilon(\mathbf{x})} [f(\mathbf{y}) - f(\mathbf{x})]\varphi_\epsilon(\mathbf{x} - \mathbf{y})d\mathbf{y}. \tag{5.20}$$

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. The fact that φ has total integral one implies that $\hat{\varphi}(0) = 1$. This gives another heuristic for understanding what happens to $\varphi_\epsilon * f$ as ϵ tends to 0. It follows from Theorem 5.1.1 that

$$\begin{aligned}\mathcal{F}(\varphi_\epsilon * f)(\boldsymbol{\xi}) &= \hat{\varphi}_\epsilon(\boldsymbol{\xi})\hat{f}(\xi) \\ &= \hat{\varphi}(\epsilon\boldsymbol{\xi})\hat{f}(\boldsymbol{\xi}).\end{aligned} \tag{5.21}$$

Thus, for each fixed $\boldsymbol{\xi}$, the limit of $\mathcal{F}(\varphi_\epsilon * f)(\boldsymbol{\xi})$ is $\hat{f}(\boldsymbol{\xi})$.

Convolution with φ_ϵ tends to average out the noise, while, at the same time, blurring the fine detail in the image. The size of ϵ determines the degree of blurring. Because both noise *and* fine detail are carried by the high frequency components, this can be understood in the Fourier representation. Since φ has integral 1 it follows that $\hat{\varphi}(0) = 1$; as $\hat{\varphi}$ is a smooth function

$$\hat{\varphi}(\epsilon\boldsymbol{\xi}) \approx 1 \text{ for } \|\boldsymbol{\xi}\| << \epsilon^{-1}.$$

Hence, for "low frequencies," that is $\|\boldsymbol{\xi}\| << \epsilon^{-1}$, the Fourier transform $\mathcal{F}(\varphi_\epsilon * f)(\boldsymbol{\xi})$ closely approximates $\hat{f}(\boldsymbol{\xi})$. On the other hand, $\hat{\varphi}(\boldsymbol{\xi})$ tends to zero rapidly as $\|\boldsymbol{\xi}\| \to \infty$ and therefore the high frequency content of f is suppressed in $\varphi_\epsilon * f$. Using convolution to suppress noise inevitably destroys fine detail.

Example 5.2.1. Consider the function of two variables f shown (as a density plot) in figure 5.6(a). The convolution of f with a smooth function is shown in figures 5.6(b-c). Near points where f is slowly varying $\varphi_\epsilon * f$ is quite similar to f. Near points where f is rapidly varying this is not the case.

(a) A reconstruction of the Shepp-Logan phantom, see figure 3.7 .

(b) The function in (a) convolved with φ_ϵ with a small ϵ.

(c) The function in (a) convolved with φ_ϵ with a large ϵ.

Figure 5.6: Convolving f reduces the noise but blurs the detail.

Remark 5.2.2. 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

$$\psi_{k,\epsilon}(x) = \epsilon^{-1}\psi_k(\frac{x}{\epsilon})$$

and f is locally integrable, then $< \psi_{k,\epsilon} * f >$ is a family of $(k-1)$-times differentiable functions, which converge, in an appropriate sense to f.

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 $f*g$ is contained in $[a-\epsilon, b+\epsilon]$.

Exercise 5.2.2. For the functions ψ_k, defined in (5.22), find the constants c_k so that

$$\int_{-1}^{1} \psi_k(x)dx = 1.$$

5.2.2 Some convergence results*

We now prove some precise results describing different ways in which $\varphi_\epsilon * f$ converges to f. For most of these results it is only necessary to assume that φ is an L^1-function with total integral one. The 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 L^1-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 at $\boldsymbol{\xi}$, computed using (4.37), is

$$\mathcal{F}(\varphi_\epsilon)(\boldsymbol{\xi}) = \hat{\varphi}(\epsilon\boldsymbol{\xi}). \tag{5.23}$$

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, (5.23) and the fact that $\hat{\varphi}(0) = 1$ imply that

$$\lim_{\epsilon\to\infty} \|\hat{f}(\hat{\varphi}_\epsilon - 1)\|_{L^2} = 0.$$

□

A similar result holds in the L^1-case.

Proposition 5.2.3. *Suppose that φ is an L^1-function with*

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

If f belongs to $L^1(\mathbb{R}^n)$ 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 belongs to $L^1(\mathbb{R}^n)$ 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

$$\begin{aligned}
\|\varphi_\epsilon * f - f\|_{L^1} &= \int_{\mathbb{R}^n}\left|\int_{\mathbb{R}^n}[f(\mathbf{x}-\epsilon\mathbf{t}) - f(\mathbf{x})]\varphi(\mathbf{t})d\mathbf{t}\right| d\mathbf{x} \\
&\leq \int_{\mathbb{R}^n}|\varphi(\mathbf{t})|\left[\int_{\mathbb{R}^n}|f(\mathbf{x}-\epsilon\mathbf{t}) - f(\mathbf{x})|d\mathbf{x}\right] d\mathbf{t} \\
&= \int_{\mathbb{R}^n}|\varphi(\mathbf{t})|\|f_{\epsilon\mathbf{t}} - f\|_{L^1}d\mathbf{t}.
\end{aligned} \tag{5.24}$$

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. □

Finally it is useful to examine $\varphi_\epsilon * f(\mathbf{x})$ at points where f is smooth. Here we use a slightly different assumption on φ.

Proposition 5.2.4. *Let f be a locally integrable function and φ an integrable function with bounded support and total integral one. 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

$$\|\mathbf{x} - \mathbf{y}\| < \delta \Rightarrow |f(\mathbf{x}) - f(\mathbf{y})| < \eta. \tag{5.25}$$

This implies that $|f(\mathbf{y})|$ is bounded for $\mathbf{y}$ in $B_\delta(\mathbf{x})$. If ϵ is sufficiently small, say less than ϵ_0, then the support of φ_ϵ is contained in the ball of radius δ and therefore $\varphi_\epsilon * f(\mathbf{x})$ is defined by an absolutely convergent integral. Since the total integral of φ is 1 we have, for an $\epsilon < \epsilon_0$, that

$$\begin{aligned} |\varphi_\epsilon * f(\mathbf{x}) - f(\mathbf{x})| &= \left| \int\limits_{B_\delta} \varphi_\epsilon(\mathbf{y})(f(\mathbf{x}-\mathbf{y}) - f(\mathbf{x})) d\mathbf{y} \right| \\ &\leq \int\limits_{B_\delta} |\varphi_\epsilon(\mathbf{y})||f(\mathbf{x}-\mathbf{y}) - f(\mathbf{x})| d\mathbf{y} \\ &\leq \int\limits_{B_\delta} |\varphi_\epsilon(\mathbf{y})| \eta d\mathbf{y} \\ &\leq \|\varphi\|_{L^1} \eta. \end{aligned} \tag{5.26}$$

In the third line we use the estimate (5.25). Since $\eta > 0$ is arbitrary this completes the proof of the proposition. $\square$

Remark 5.2.3. There are many variants of these results. The main point of the proofs 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.

We close this section by applying the approximation results above to 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

$$F(\mathbf{x}) = \int\limits_{\mathbb{R}^n} \hat{f}(\boldsymbol{\xi})^{i\langle \mathbf{x}, \boldsymbol{\xi}\rangle} d\boldsymbol{\xi}.$$

Proposition 5.2.3 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. $\square$

Exercises

Exercise 5.2.3. Use Corollary A.7.1 to prove Lemma 5.2.1.

Exercise 5.2.4. Give the details of the argument, using Lemma 5.2.1, to show that if f is an L^1-function, then

$$\lim_{\epsilon\to 0}\int_{\mathbb{R}^n} \varphi(\mathbf{t})\|f_{\epsilon\mathbf{t}} - f\|_{L^1} d\mathbf{t} = 0.$$

Exercise 5.2.5. Use the method used to prove Proposition 5.2.4 to show that if $f \in L^p(\mathbb{R})$ for a $1 \le p < \infty$ 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 5.2.6. Let $\psi_\epsilon(x) = [2\epsilon]^{-1}\chi_{[-\epsilon,\epsilon]}(x)$. Show by direct computation that if $f \in L^2(\mathbb{R})$ then $\psi_\epsilon * f$ converges to f in $L^2(\mathbb{R})$.

5.2.3 Approximating derivatives and regularized derivatives

If either f or φ is a differentiable function then $\varphi * f$ is as well. In this section we assume that φ is a bounded function with support in B_1 and total integral one. If f has k continuous derivatives in $B_\delta(\mathbf{x})$ then, for $\epsilon < \delta$ the convolution $\varphi_\epsilon * f$ is k-times differentiable. For each $\boldsymbol{\alpha}$ with $|\boldsymbol{\alpha}| \le k$, Proposition 5.2.1 implies that

$$\partial_{\mathbf{x}}^{\alpha}(\varphi_\epsilon * f)(\mathbf{x}) = \varphi_\epsilon * \partial_{\mathbf{x}}^{\boldsymbol{\alpha}} f(\mathbf{x}).$$

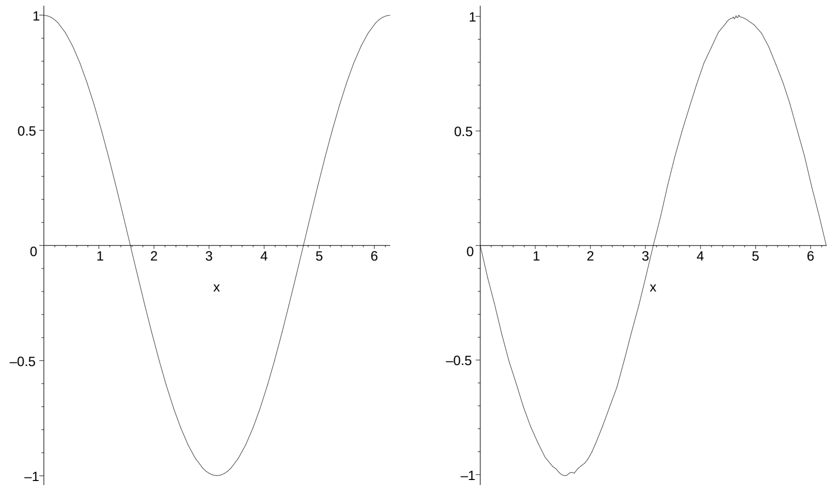
Proposition 5.2.4 can be applied to conclude that

$$\lim_{\epsilon\to 0}\partial_{\mathbf{x}}^{\alpha}(\varphi_\epsilon * f)(\mathbf{x}) = \partial_{\mathbf{x}}^{\boldsymbol{\alpha}} f(\mathbf{x}).$$

On the other hand, if f is not a differentiable function but φ is, then

$$\partial_{\mathbf{x}_j}(\varphi_\epsilon * f) = (\partial_{\mathbf{x}_j}\varphi_\epsilon) * f$$

can be used to *define* regularized approximations to the partial derivatives of f. This can be useful if f is the result of a noisy measurement of a smooth function which, for one reason or another, must be differentiated. Precisely this situation arises in the reconstruction process used in x-ray CT-imaging. We illustrate this idea with an example:



(a) A noisy function smoothed by convolution with a triangle function, $\epsilon = .1$.

(b) The regularized derivative of the noisy function, $\epsilon = .1$.

Figure 5.7: Using convolution to find a regularized approximate derivative.

Example 5.2.2. Let f be the noise corrupted version of $\cos(x)$ depicted in figure 5.1(c). To smooth f we use the "triangle function"

$$t_\epsilon(x) = \begin{cases} \frac{\epsilon - |x|}{\epsilon^2} & \text{if } |x| \leq \epsilon, \\ 0 & \text{if } |x| > \epsilon. \end{cases}$$

The derivative of $f * t_\epsilon$ is computed using the weak derivative of t_ϵ. The result of computing $f * t_{.1}$ is shown in figure 5.7(a) while $f * t'_{.1}$ is shown in figure 5.7(b). The approximation to $-\sin(x)$ provided by $f * t'_{.1}$ is impressive given that $t_{.1}$ has only a weak derivative.

Exercise

Exercise 5.2.7. * 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.

5.2.4 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 5.2.2. 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.

Proposition 5.2.5. *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

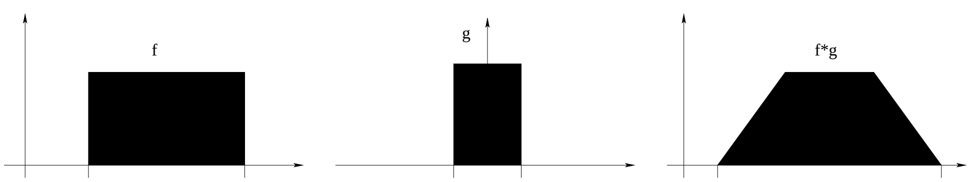
$$f * g(\mathbf{x}) = \int\limits_{\mathbb{R}^n} f(\mathbf{y})g(\mathbf{x}-\mathbf{y})d\mathbf{y} = 0$$

as well. Because $\operatorname{supp} f + \operatorname{supp} g$ is a closed set there is an $\eta > 0$ such that $B_\eta(\mathbf{x})$ is disjoint from $\operatorname{supp} f + \operatorname{supp} g$. The argument showing that $f * g(\mathbf{x})$ equals 0 applies to any point $\mathbf{x}'$ in $B_\eta(\mathbf{x})$ and therefore proves the proposition. □

If φ is a function supported in the ball of radius one then φ_ϵ is supported in the ball of radius ϵ. According to Proposition 5.2.5 the support of $\varphi_\epsilon * f$ is contained in the set

$$\{\mathbf{x} + \mathbf{y} \ : \ \mathbf{x} \in \operatorname{supp} f \text{ and } \mathbf{y} \in B_\epsilon\}.$$

These are precisely the points that are within distance ϵ of the support of f, giving another sense in which ϵ reflects the resolution available in $\varphi_\epsilon * f$. Figure 5.8 shows a one dimensional example.



(a) $\operatorname{supp} f$ (b) $\operatorname{supp} g$ (c) $\operatorname{supp} f + \operatorname{supp} g$

Figure 5.8: The support of $f * g$ is contained in $\operatorname{supp} f + \operatorname{supp} g$.

Example 5.2.3. 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

$$\begin{aligned} \int\limits_{-\infty}^{\infty} \psi * \psi(x)dx &= \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} \psi(y)\psi(x-y)dydx \\ &= \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} \psi(y)\psi(t)dtdy \\ &= 1 \cdot 1 = 1. \end{aligned} \tag{5.27}$$

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.

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

$$f(0) = \int_{\mathbb{R}^n} \delta(\mathbf{y}) f(\mathbf{y}) d\mathbf{y}. \tag{5.28}$$

Proceeding formally we see that

$$\begin{aligned} \delta * f(\mathbf{x}) &= \int_{\mathbb{R}^n} \delta(\mathbf{y}) f(\mathbf{x} - \mathbf{y}) d\mathbf{y} \\ &= f(\mathbf{x} - \mathbf{0}) = f(\mathbf{x}). \end{aligned} \tag{5.29}$$

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 (5.28) only makes sense for functions continuous at 0, the convolution of δ with an arbitrary locally integrable function 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 (5.18) to approximate δ in $\mathbf{x}$-space. The other is to approximate $\hat{\delta}$ in $\boldsymbol{\xi}$-space. To close this chapter we consider some practical aspects of approximating the δ-function in one-dimension and formalize the concept of resolution.

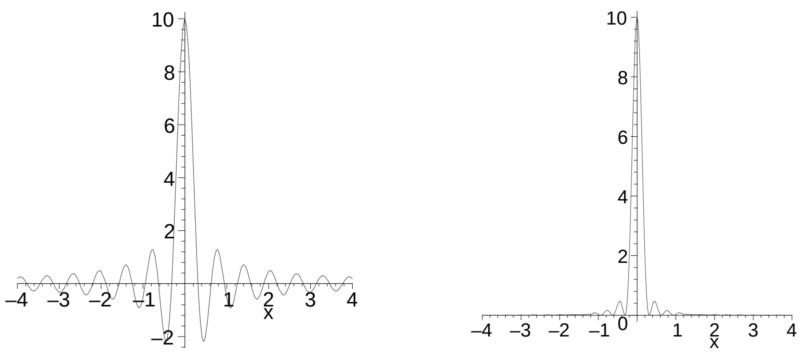
5.3.1 Approximating the δ-function in 1-dimension

Suppose that φ is an even function with bounded support and total integral one. 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, $\partial_\xi \hat{\varphi}(0) = 0$; thism means that the behavior of $\hat{\varphi}(\xi)$ for ξ near to zero is largely governed by the "second moment"

$$\partial_\xi^2 \hat{\varphi}(0) = -\int\limits_{-\infty}^{\infty} x^2 \varphi(x) dx.$$

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 5.9.

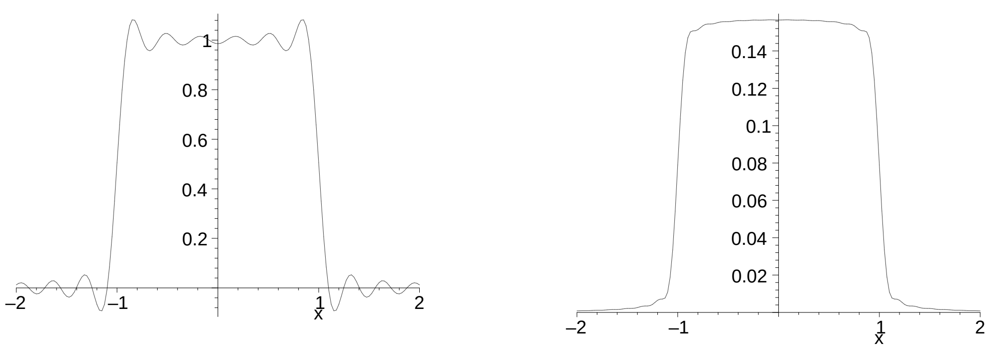


(a) A sinc-pulse.

(b) A sinc^2-pulse.

Figure 5.9: 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 5.10. In figure 5.10(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 5.10: 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 5.5. 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.

(5.30)

Definition 5.3.1. Let ψ satisfy the conditions in (5.30) 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 $\text{FWHM}(\psi)$.

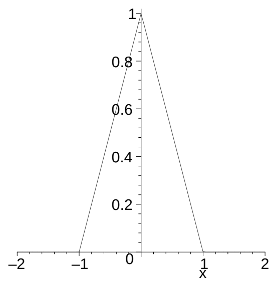
If f is an input then the resolution available in the output, $\psi * f$ is *defined* to be the $\text{FWHM}(\psi)$. In principle if $\text{FWHM}(\psi_1) < \text{FWHM}(\psi_2)$ then $f \mapsto \psi_1 * f$ should have better resolution than $f \mapsto \psi_2 * f$. Here is a heuristic explanation for this definition. Suppose that the signal f is pair of unit impulses separated by a distance d,

$$f(x) = \delta(x) + \delta(x-d).$$

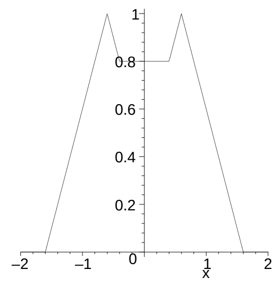
Convolving ψ with f produces two copies of ψ,

$$\psi * f(x) = \psi(x) + \psi(x-d).$$

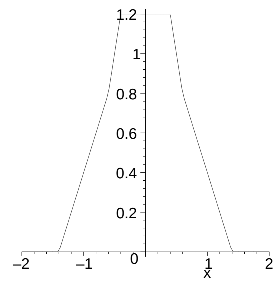
If $d > \text{FWHM}(\psi)$ then $\psi * f$ has two distinct maxima separated by a valley. If $d \leq \text{FWHM}(\psi)$ then the distinct maxima disappear. If the distance between the impulses is greater than the $\text{FWHM}(\psi)$ then we can "resolve" them in the filtered output. More generally the $\text{FWHM}(\psi)$ is considered to be the smallest distance between distinct "features" in f which can be seen in $\psi * f$. In figure 5.11 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 $.8 < 1$.



(a) $\psi(x)$



(b) $\psi(x+.6)+\psi(x-.6)$



(c) $\psi(x+.4)+\psi(x-.4)$

Figure 5.11: Illustration of the FWHM definition of resolution

This FWHM-definition of resolution is often applied to filters defined by functions which do not satisfy all the conditions in (5.30) 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 $f \mapsto f * \operatorname{sinc}$. The FWHM(sinc) is taken to be the full width half maximum of its central peak, it is approximately given by

$$\text{FWHM}(\operatorname{sinc}) \approx 3.790988534.$$

We return to the problem of quantifying resolution in Chapter 9.

Exercises

Exercise 5.3.3. Numerically compute the FWHM$(\operatorname{sinc}^2(x))$. How does it compare to FWHM$(\operatorname{sinc}(x))$.

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,

$$\text{FWHM}(h_j) \simeq 2\sqrt{\frac{6\log 2}{j}}.$$

Exercise 5.3.5. Using the Taylor expansion for the sine, show that as B gets large

$$\text{FWHM}(\operatorname{sinc}(Bx)) \simeq \frac{2\sqrt{3}}{B}.$$

Exercise 5.3.6. For $a > 0$ let $g_a(x) = e^{-\frac{x^2}{a^2}}$. Compute FWHM$(g_a)$ and FWHM$(g_a * g_b)$.