1 A New Kind of Product of Functions

You are familiar with the pointwise product of functions defined by \( f \cdot g(x) = f(x) \cdot g(x) \). You just take the product of the real numbers \( f(x) \) and \( g(x) \). Thus if we define \( f(x) = 1 \) for all \( x \), we get \( f \cdot g(x) = g(x) \). So \( f(x) = 1 \) is the identity for pointwise product.

Now we want to define a new kind of product, well, not that new if you got to convolution in the Laplace transform section of 20D. And not so new if you are taking probability or statistics courses where given independent random variables with densities \( f \) and \( g \), the density of the sum of the random variables is \( f \cdot g \).

Our aim is to use convolution in order to uniformly approximate continuous functions \( f \) on \([a, b]\) by polynomials \( \sum_{n=0}^{N} a_n x^n \) (or functions of period 1 by trigonometric polynomials \( \sum_{n=-\infty}^{\infty} a_n e^{2\pi i nx} \) for Fourier series).

We will assume that our functions are piecewise continuous and that at least one of the functions in \( f \cdot g \) vanishes off a bounded interval so that we know the integrals exist.

**Definition 1** The convolution of \( f \) and \( g \) is defined for all \( x \) by

\[
(f \ast g)(x) = \int_{-\infty}^{\infty} f(t)g(x-t)dt.
\]

One can read \( f \ast g \) this as \( f \) splat \( g \) since it does splat the properties of the 2 functions together, preserving the best properties. If one function is discontinuous, but the other is a polynomial, the result is a polynomial.

**Example.** Let \( g(x) = 1 - x^2 \), for all \( x \). Define

\[
f(x) = \begin{cases} 
1, & \text{if } 0 < x < 1 \\
-1, & \text{if } -1 < x < 0 \\
0, & \text{if } x = 0 \\
0, & \text{if } x \geq 1 \text{ or } x \leq -1.
\end{cases}
\]

Then, since \( f \) has differing formulas on different intervals (and mercifully \( g \) does not), we get

\[
(f \ast g)(x) = -1 \int_{-1}^{0} g(x-t)dt + 1 \int_{0}^{1} g(x-t)dt
\]

\[
= -\int_{-1}^{0} (1 - (x-t)^2) dt + \int_{0}^{1} (1 - (x-t)^2) dt
\]

\[
= \int_{-1}^{0} (x-t)^2 dt - \int_{0}^{1} (x-t)^2 dt
\]

\[
= \frac{-(x-t)^3}{3} \bigg|_{-1}^{0} - \frac{-(x-t)^3}{3} \bigg|_{0}^{1}
\]

\[
= \frac{-2x^3}{3} + \frac{(x+1)^3}{3} + \frac{(x-1)^3}{3}
\]

\[
= \frac{3x^2 + 3x - 3x^2 + 3x}{3} = 2x
\]
Figure 1: \( g(x) = 1 - x^2 \)

Figure 2: The function \( f(x) \)
So we see that even though \( f \) is discontinuous, when convolved with a polynomial, we get a polynomial.

**Properties of Convolution.**

Assume \( f, g, h \) are piecewise continuous and at least one vanishes off a bounded interval when necessary to make an integral exist. Let \( \alpha \in \mathbb{R} \).

1) \( f \ast g = g \ast f \)
2) \( f \ast (g + h) = f \ast g + f \ast h \)
3) \( (\alpha f) \ast g = \alpha (f \ast g) \)
4) \( f \ast (g \ast h) = (f \ast g) \ast h \)
5) Suppose that \( f(x) = 0 \) if \( |x| > c \) and \( g(x) = 0 \) if \( |x| > d \). Then \( (f \ast g)(x) = 0 \) if \( |x| > c + d \).
6) If \( g(x) \) is a polynomial then \( (f \ast g)(x) \) is a polynomial.
7) \( (f \ast g) = f \ast (g') \) assuming \( g \) differentiable.

**Proof.**

1) Make the change of variables \( u = x - t \). Then \( t = x - u \) and \( du = -dt \) and the order of integration changes so we get:

\[
f \ast g(x) = \int_{-\infty}^{\infty} f(t)g(x-t)dt = -\int_{-\infty}^{-c} f(x-u)g(u)du + \int_{c}^{\infty} g(u)f(x-u)du = g \ast f(x).
\]

2)-4) are Exercises in homework 7.

5) \[
\int_{-\infty}^{\infty} f(t)g(x-t)dt = \int_{-c}^{c} f(t)g(x-t)dt.
\]

Now if \( x > c + d \), and \( -c \leq t \leq c \), we see that \( x - t > c + d - c = d \) and \( g(x-t) = 0 \). So \( x > c + d \) implies \( (f \ast g) = 0 \).

If \( x < -c - d \) and \( -c \leq t \leq c \), we see that \( x - t < -c - d + c = -d \) and \( g(x-t) = 0 \). Thus \( x < -c - d \) implies \( (f \ast g) = 0 \).

6) It suffices using 2) and 3) to consider the case that \( g(x) = x^n \). Then supposing \( f(x) = 0 \) if \( |x| > c \), we have the following, using the binomial theorem

\[
f \ast g(x) = \int_{-\infty}^{\infty} f(t)g(x-t)dt = \int_{-c}^{c} f(t)(x-t)^n dt
= \int_{-c}^{c} f(t)\sum_{k=0}^{n} \binom{n}{k} x^{n-k}(-t)^k dt = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} \int_{-c}^{c} f(t)(-t)^k dt.
\]
Let
\[ \int_{-c}^{c} f(t)(-t)^{n-k} dt = c_k. \]

Then
\[ f \ast g(x) = \sum_{k=0}^{n} \binom{n}{k} c_k x^{n-k}, \]
which is a polynomial.

7) We will not need this so we leave it as an extra credit exercise. ■

This concludes our discussion of convolution. It is important in the theory of probability and Fourier analysis, Laplace transforms. In fact the Fourier and Laplace transforms change convolution into ordinary pointwise product of the transformed functions. Convolution is also used to smooth data thanks to property 7). We want to use it to approximate continuous functions by polynomials.

## 2 A Function Which is Not an Ordinary Function

We want to think about something called the Dirac delta "function" denoted \( \delta(x) \). It is used in physics to represent an impulse. It is often said to be a function that is 0 for \( x \neq 0 \) and \( \infty \) at \( x = 0 \). Of course, no mathematician would call that legal. The graph usually associated with \( \delta \) shows a unit arrow or spike at the origin - not a point at infinite height.

Another "definition" of \( \delta \) is that for any continuous function \( f(x) \) it is supposed to give the formula
\[ \int_{-\infty}^{\infty} f(t)\delta(t)dt = f(0). \]

We show in the homework that this formula forces \( \delta(x) = 0 \) for \( x \neq 0 \) and thus forces the integral to be 0.

Yet another definition of delta is that it is the identity for convolution. That is just as bad.
This disturbed mathematicians mightily in the 1930’s and 1940’s. Engineers and physicists just happily used delta and even its derivative or infinite sums of deltas. Finally Laurent Schwartz and others legalized delta and its relatives by calling it a distribution or generalized function and developing the calculus of distributions. See my book Harmonic Analysis on Symmetric Spaces and Applications, I, for an introduction. Other references are Korevaar, Mathematical Methods and Schwartz, Math. for the Physical Sciences. The subject of analysis on distributions does not seem to appear in undergrad math courses.

So what to do? In this section we consider sequences of functions that have the behavior of delta in the limit as $n \to \infty$. These are called Dirac sequences or approximate identities (for convolution).

**Definition 2** A Dirac sequence $K_n(x)$ is a sequence of functions $K_n : \mathbb{R} \to \mathbb{R}$ such that

Dir 1. $K_n(x) \geq 0$, for all $x$ and $n$.

Dir 2. $\int_{-\infty}^{\infty} K_n(x) dx = 1$, for all $n$.

Dir 3. $\forall \varepsilon > 0$ and $\forall \delta > 0$, $\exists \ N \in \mathbb{Z}^+$ s.t. $n \geq N$ implies

$$\int_{-\infty}^{\infty} K_n(x) dx + \int_{\delta}^{\infty} K_n(x) dx < \varepsilon.$$

The 3 properties say that the area under the curve $y = K_n(x)$ becomes more and more concentrated at the origin as $n \to \infty$. You might think it hard to find such a sequence but we have the following example from the first problem in Lang.

**Example 1.** Let $K(x)$ be such that $K(x) \geq 0$ for all $x$ and $\int_{-\infty}^{\infty} K(x) dx = 1$. Then $K_n(x) = nK(nx)$ is a Dirac sequence.

**Example 2.** The Landau Kernel.

This kernel is named for a mathematician (number theorist) who lived at the same time as Dirac.

Define the Landau kernel to be

$$L_n(x) = \begin{cases} \frac{(1-x)^n}{c_n}, & -1 \leq t \leq 1 \\ 0, & \text{otherwise.} \end{cases}$$

where $c_n = \int_{-1}^{1} (1-t^2)^n dt$.

It is possible to use integration by parts to find that

$$c_n = \frac{(n!)^2 2^{2n+1}}{(2n)!(2n+1)}.$$

See Courant and Hilbert, Methods of Mathematical Physics, Vol. I, p. 84. Unless you know Stirling’s formula, this is not too helpful in proving the Landau kernel gives a Dirac sequence. Lang gives a simple inequality which is all we need. In Figure 5, we plot $L_n(t) = (1-t^2)^n \frac{(2n)!(2n+1)}{(n!)^2 2^{2n+1}}$ for $t \in [-1, 1]$, when $n = 10$ in blue; 50 in green; 100 in turquoise; and 200 in purple. Figure 5 shows the area under the curve (which is 1) begins to concentrate at the origin as $n$ increases.

**Lemma.** $c_n \geq \frac{2}{n+1}$.

**Proof.**

$$\frac{c_n}{2} = \int_{0}^{1} (1-t^2)^n dt \geq \int_{0}^{1} (1-t)^n (1+t)^n dt = \frac{1}{n+1}.$$

It is clear that the Landau kernel has the first 2 properties of a Dirac sequence. To prove the 3rd property, we argue as in Lang. Given $\varepsilon > 0$ and $0 < \delta < 1$, we need to find $N$ so that $n \geq N$ makes the following integral $< \varepsilon$:

$$\frac{1}{c_n} \int_{\delta}^{1} (1-t^2)^n dt \leq \frac{n+1}{2} \int_{\delta}^{1} (1-t^2)^n dt \leq \frac{n+1}{2} \int_{\delta}^{1} (1-\delta^2)^n dt = \frac{n+1}{2} (1-\delta^2)^n (1-\delta).$$

We can make the stuff on the right $< \varepsilon$ for large $n$, since we can show that because $0 < 1-\delta^2 < 1$, $\lim_{n \to \infty} \frac{n+1}{2} (1-\delta^2)^n = 0$. Use l’Hôpital’s rule.
3 A Dirac Sequence Approaches The Dirac Delta

We want to prove that any Dirac sequence $K_n$ behaves like an identity for convolution in the limit as $n \to \infty$. Some people call Dirac sequences "approximate identities" for this reason.

**Theorem 3** Suppose $f$ is a bounded piecewise continuous function on $\mathbb{R}$. Let $I$ be any finite interval on which $f$ is continuous. Define $\|g\|_\infty = \max_{x \in I} |g(x)|$. Then $\lim_{n \to \infty} \|f - K_n \ast f\|_\infty = 0$. This says that the sequence $K_n \ast f$ converges uniformly to $f$ on the interval $I$.

**Proof.** Using the 1st and 2nd properties of Dirac sequences plus the fact that integrals preserve $\leq$, we have:

\[
\left| \int K_n(t) f(x-t) dt - f(x) \right| = \left| \int K_n(t) f(x-t) dt - f(x) \int K_n(t) dt \right|.
\]

\[
= \left| \int K_n(t) (f(x-t) - f(x)) dt \right|.
\]

\[
\leq \int |K_n(t)| |f(x-t) - f(x)| dt.
\]

Since $f$ is uniformly continuous on $I$, given $\varepsilon$ there is a $\delta$ such that $|f(x-t) - f(x)| < \varepsilon$ when $|t| < \delta$. Since $f$ is bounded, we there is a bound $M$ so that $|f(x)| \leq M$ for all $x$.

Now we can break up the last integral

\[
\int_{-\infty}^{\infty} |K_n(t)| |f(x-t) - f(x)| dt = \int_{-\infty}^{-\delta} |K_n(t)| |f(x-t) - f(x)| dt + \int_{-\delta}^{\infty} |K_n(t)| |f(x-t) - f(x)| dt + \int_{-\delta}^{\delta} |K_n(t)| |f(x-t) - f(x)| dt.
\]

For the first 2 integrals, use the bound on $f$ and the 3rd property of Dirac sequences to see that, for large enough $n$, they are less than $2M\varepsilon$ or even $\varepsilon$, if you prefer.
For the last integral, use the uniform continuity of $f$ to see that the integral is

$$
\leq \varepsilon \int_{-\delta}^{\delta} K_n(t)dt \leq \varepsilon \int_{-\infty}^{\infty} K_n(t)dt = \varepsilon.
$$

This completes the proof that $|(K_n * f)(x) - f(x)| \leq (2M + 1)\varepsilon$. You can replace $\varepsilon$ by $\frac{\varepsilon}{2M+1}$ if you are paranoid. 

**Corollary.** **Weierstrass Theorem.** Any continuous function $f$ on a finite closed interval $[a, b]$ can be uniformly approximated by polynomials.

**Proof.** We can use the preceding theorem for the interval $[0, 1]$ along with the Landau kernel. See Lang, p. 288 for a general interval. We still need to see that $L_n * f$ is a polynomial. Suppose $x \in [0, 1]$. Then

$$
\int_{-\infty}^{\infty} L_n(x-t)f(t)dt = \int_{0}^{1} L_n(x-t)f(t)dt.
$$

Also we see that for $x, t \in [0, t]$, we have $-1 \leq x-t \leq 1$ and so $L_n(x-t) = \left(1 - (x-t)^2\right)^n$ and the integral is

$$
\frac{1}{c_n} \int_{0}^{1} \left(1 - (x-t)^2\right)^n f(t)dt.
$$

This is a polynomial using the same reasoning as in our proof in the 1st section that convolution of $f$ with any polynomial is a polynomial. 

This finishes our story for the moment.