Assigning a topology to the Schwartz class

The Schwartz class is interesting, and applicable, with the Fourier transform and the tools of Lebesgue measure. We will define open sets for S, and with this topology, introduce the space S' of continuous linear functionals on S. These objects are also called **tempered distributions**. They "include" all the L^p spaces, S itself, and much more besides. The key ideas are these: regarding the action of a linear functional as a "formal" integral, and the "transpose" of a linear mapping. We have seen this in two places recently: the equations $\int \frac{\partial f}{\partial x_k} g\, dx = -\int f \frac{\partial g}{\partial x_k} dx$ for integration by parts, and $\int \hat{f}(\xi)\, g(\xi)\, d\xi = \int f(x)\, \hat{g}(x)\, dx$, that says we can "move the hat." The first equation is valid with f in S and g in M. In that equation, the left side makes sense for *any* measurable function g bounded by a polynomial, if f is in S. So we *define* a linear functional on S by the equation $T_g(f) := \int f(x)\, g(x)\, dx$. The left side of the first equation can then be expressed as $T_g(\frac{\partial f}{\partial x_k})$. If g is in M, then we can write the equation as $T_g(\frac{\partial f}{\partial x_k}) = -T_{\partial g/\partial x_k}(f)$. The right side may *not* make sense, if g is not in M, so we *define* $\frac{\partial g}{\partial x_k}$ to be the *linear functional*, defined on S, given by the left side. The same is true in the second equation. The second equation is valid with f in S and g in S. The left side makes sense if f is in S and g is in M. But the right side does not — how do we define the Fourier transform of $|x|^2$, for example? The point is, we define it by the equation, not as a function, but as a *linear functional*. In fact, we have $\hat{f}(\xi)\ \xi_k{}^2 = -\hat{f}(\xi)\ (i\xi_k)^2 = -\left(\frac{\partial^2 f}{\partial x_k{}^2}\right)^\wedge(\xi)$, so $\hat{f}(\xi)\ |\xi|^2 = -(\Delta f)^\wedge(\xi)$, where Δ denotes the Laplace operator. Now the equation reads $-\int (\Delta f)^\wedge(\xi)\, d\xi = (|x|^2)^\wedge(f)$. The left side contains the factor $1 = e^{-i\xi \bullet 0}$. By the inversion theorem we recognize this as $-(2\pi)^n(\Delta f)(0)$. Thus we will be able to say $(|x|^2)^\wedge = -(2\pi)^n(\Delta\bullet)(0)$.

two possible topologies for S

We have infinitely many norms on S, namely the $\| f \|_{\alpha\beta}$, with α and β in $\mathbf{N}^n$. Each of these gives us a collection of open sets, i.e. a topology on S. There are two immediate possibilities: we can define our topology to be the intersection of all these norm topologies. This will certainly yield a topology. In fact, we can use this fact to talk about the smallest topology containing a given collection of sets. The other possibility is the one **we will use: the smallest topology that contains *all* these norm topologies**. It is larger than the first one. Thus it is "easier" for a function that has S as its domain to be continuous. Let X be a set that has a topology. Then,
f: $S \to X$ is continuous with respect to our first topology if and only if f is continuous with respect to every topology in the collection. On the other hand, f: $S \to X$ is continuous with respect to our second topology if f is continuous with respect to just one topology in the collection.

Theorem: The smallest topology on S that contains all the topologies $T_{\alpha\beta}$ given by the norms $\| f \|_{\alpha\beta}$, with α and β in $\mathbf{N}^n$, is the collection of arbitrary unions of finite intersections of members of $\bigcup_{\alpha\beta} \mathcal{T}_{\alpha\beta}$.

Proof: Any topology that contains $\bigcup_{\alpha\beta} \mathcal{T}_{\alpha\beta}$ certainly contains finite intersections from this union, and arbitrary unions of them. And this collection of arbitrary unions of finite intersections is a topology on $\mathcal{S}$. Thus it is the desired topology.

how to use this topology

Lemma: Let U be a subset of $\mathcal{S}$. Then U is open if and only if, for all f in U, there exists a finite subset Φ of $\mathbf{N}^{n2} := \mathbf{N}^n \times \mathbf{N}^n$, and positive numbers $\varepsilon_{\alpha\beta}$, such that $\| f - g \|_{\alpha\beta} < \varepsilon_{\alpha\beta}$ for all (α,β) in Φ implies that $g \in U$.

Proof: Suppose U is open. Then for each f in U there exists a set Ω, contained in U, with $f \in \Omega$, that is the union of sets that are finite intersections of sets in $\bigcup_{\alpha\beta} \mathcal{T}_{\alpha\beta}$. Now, f belongs to one of these finite intersections. This means that there are finitely many open sets, each belonging to one of the topologies $\mathcal{T}_{\alpha\beta}$, such that f is in each one. In order that f belong to a set $\mathcal{O}$ that is open in $\mathcal{T}_{\gamma\delta}$, there must be a positive number $\varepsilon_{\gamma\delta}$ such that $B_{\gamma\delta}(f,\varepsilon_{\gamma\delta}) \subseteq \mathcal{O}$. That is, for all g in $\mathcal{S}$, $\| f - g \|_{\gamma\delta} < \varepsilon_{\gamma\delta}$ implies that $g \in U$. It follows that the set Φ and positive numbers $\varepsilon_{\alpha\beta}$ asserted by the lemma exist.

Next, suppose the set Φ and positive numbers $\varepsilon_{\alpha\beta}$ mentioned in the lemma exist, for each f in a set U contained in $\mathcal{S}$. Then each ball $B_{\alpha\beta}(f,\varepsilon_{\alpha\beta})$ is open, and contained in U. Thus U is open.

Corollary: If a subset U of $\mathcal{S}$ is a neighborhood of 0, there exists a finite subset Φ of $\mathbf{N}^{n2}$, and positive numbers $\varepsilon_{\alpha\beta}$, such that $\| f \|_{\alpha\beta} < \varepsilon_{\alpha\beta}$ for all (α,β) in Φ implies that $f \in U$. Conversely, the set of all f in $\mathcal{S}$ for which such inequalities are true is a neighborhood of 0.

Proof: The first statement is, by definition, the statement that 0 is an interior point of U. The second statement follows from the fact that the indicated “neighborhood” is the intersection of open sets that each contain 0.

Lemma: A linear mapping $T: \mathcal{S} \to V$, where V is a normed space, is continuous if and only if there exists a finite subset Φ of $\mathbf{N}^{n2}$, and positive numbers $C_{\alpha\beta}$, such that for all f in $\mathcal{S}$,

$$\| Tf \| \leq \sum_{(\alpha,\beta)\in\Phi} C_{\alpha\beta} \| f \|_{\alpha\beta}.$$

Proof: Here, $\| Tf \|$ is the norm of Tf in V. Because our topology consists of all translates of open sets containing 0, and because T is linear, to prove continuity we only have to prove continuity at 0. Conversely, given the continuity, we know we have continuity at 0. Suppose the given norm inequality holds. Then if, for any $\varepsilon > 0$, we consider those f such that

$\| f \|_{\alpha\beta} < \dfrac{\varepsilon}{\sum_{\gamma\delta\in\Phi} C_{\gamma\delta}}$ for each $(\alpha,\beta) \in \Phi$, we have a neighborhood of 0 in which $\| Tf \| < \varepsilon$. Thus T is continuous at 0, hence continuous.

Now we suppose that T is continuous at 0, and we want to show that an inequality like the one in the lemma holds. What we are given is that for all $\varepsilon > 0$, there exists a neighborhood U of 0 such that $\| Tf \| < \varepsilon$ for all f in U. Thus there exists a finite subset Φ of $\mathbf{N}^{n2}$, and positive numbers $\varepsilon_{\alpha\beta}$, such that $\| f \|_{\alpha\beta} < \varepsilon_{\alpha\beta}$ for all (α,β) in Φ implies that $f \in U$. Let B be the subset of U in which these inequalities are true.

Then $f \in B$ implies $\| Tf \| < \varepsilon$. Let f be an arbitrary non-zero element of S. Define $N(f) := \sum_{\gamma\delta\in\Phi} \frac{\| f \|_{\gamma\delta}}{\varepsilon_{\gamma\delta}}$. Then $g := \frac{f}{N(f)}$ satisfies $\| g \|_{\alpha\beta} < \varepsilon_{\alpha\beta}$ for each $(\alpha,\beta) \in \Phi$. Thus, $g \in B$, and this implies $\| Tg \| < \varepsilon$, or, $\left\| T\left(\frac{f}{N(f)}\right)\right\| < \varepsilon$. That is, $\| Tf \| \le \varepsilon \sum_{\gamma\delta\in\Phi} \frac{\| f \|_{\gamma\delta}}{\varepsilon_{\gamma\delta}}$, an inequality that also holds when $f = 0$. This is what we wanted to show.

As an application, recall the proof that $S \subseteq L^1$. We used the fact that $(1 + |x|^2)^n |f(x)|$ is bounded, by definition of S, to conclude that

$$\| f \|_1 \le \sup_x \{(1 + |x|^2)^n |f(x)|\} \int (1 + |y|^2)^{-n} \, dy.$$

We have $(1 + |x|^2)^n = \sum_{|\alpha|\le n} C_\alpha x^{2\alpha}$, for some constants C_α. Thus

$$\sup_x \{(1 + |x|^2)^n |f(x)|\} \le \sum_{|\alpha|\le n} C_\alpha \| f \|_{\alpha 0}, \text{ so } \| f \|_1 \le \sum_{|\alpha|\le n} B_\alpha \| f \|_{\alpha 0},$$

where $B_\alpha = C_\alpha \int (1 + |y|^2)^{-n} \, dy$.

Thus the “inclusion mapping,” that takes f in S into f in L^1, is continuous. A similar argument shows the continuity of the inclusion mappings of S into L^p, $1 < p < \infty$, are continuous. The inclusion mapping of S into L^∞ is immediately continuous, because the norm $\| f \|_{00}$ is the same as the L^∞ norm.

Exercise: Prove that if V is a normed space, and $T: V \to S$ is linear, then T is continuous if and only if for all (α,β) in $\mathbf{N}^{n2}$, there exists $C_{\alpha\beta}$ such that $\| Tv \|_{\alpha\beta} \le C_{\alpha\beta} \| v \|$.

Exercise: Prove that if $T: S \to S$ is linear, then T is continuous if and only if for all (α,β) in $\mathbf{N}^{n2}$, there exists a finite subset $\Phi_{\alpha\beta}$ of $\mathbf{N}^{n2}$, and positive numbers $C^{\alpha\beta}_{\gamma\delta}$, such that for all f in S, $\| Tf \|_{\alpha\beta} \le \sum_{(\gamma,\delta)\in\Phi} C^{\alpha\beta}_{\gamma\delta} \| f \|_{\alpha\beta}$.

Definition of S'

Definition: Let S' denote the collection of continuous linear functionals on S, the elements of which are called **tempered distributions**.

We will denote the generic tempered distribution by u, and write u(f) or (u, f) for the value of u at f in S. Sometimes we will write u(f) as $\int u(x) f(x)\, dx$, when u(x) is a suitable locally integrable function.

We may also write $u(f) = \int u(x) f(x)\, dx$, even when it does not make sense! This is a device useful for defining operations on distributions by extending them, by continuity, from the set of those distributions on which the operation *does* make sense. We used this device before to define $(|x|^2)^\wedge$, for example.

Examples

1. For x_o in $\mathbf{R}^n$, let $ev_{x_o}(f) := f(x_o)$. We have $|ev_{x_o}(f)| \le \| f \|_{00}$. This distribution is also denoted $\delta_{x_o}(f)$, and called the **Dirac delta function**. Another common notation for $\delta_{x_o}(f)$ is $\int \delta(x_o - x) f(x)\, dx$, where δ is thought of as a "point mass" at the origin, or as "a function that is 0 away from 0, and so large at 0 that its integral is 1." Note the use of convolution notation!

2. If u is a locally integrable function such that $(1 + |x|^2)^{-p} |u(x)|$ is integrable for some p, then $u(f) := \int u(x) f(x)\, dx$ is defined as an absolutely convergent integral for each f in S. If p is negative or 0, u(x) is integrable, and u(f) is continuous with respect to the norm $\| f \|_{00}$. If p is positive, let P be the smallest integer at least p. Then we prove the continuity of u in the same way we proved the continuity of the inclusion mapping that takes S into L^1.

3. If μ is a Borel measure such that $\int (1 + |x|^2)^{-p}\, d\mu(x) < \infty$ for some p, and g(x) is bounded and Borel measurable, then $u(f) := \int f(x) g(x)\, d\mu(x)$ is a tempered distribution, by the same argument as in the previous example.

4. We can combine these examples with continuous mappings of S into S, such as the Fourier transformation (not yet shown to be continuous!) to get examples such as $u(f) := \hat{f}(x_o)$ and $u(f) := \int \hat{f}(x)\, g(x)\, d\mu(x)$. These will be used to define the Fourier transforms of the Dirac delta and of certain measures.

5. Another version of the previous example involves differentiation, and multiplication by polynomials: $u(f) := \int x^\alpha f^{(\beta)}(x)\, dx$, $u(f) := \int x^\alpha f^{(\beta)}(x)\, g(x)\, d\mu(x)$ are tempered distributions. To verify this we need only show that the map that sends f to $x^\alpha f^{(\beta)}(x)$ is continuous from S to S. To do so, we need to estimate the $\gamma\delta$-norm of $x^\alpha f^{(\beta)}(x)$. This means we want to find the supremum of the absolute value of $x^\gamma D^\delta(x^\alpha f^{(\beta)}(x))$ = (by Leibniz' Rule) a sum of terms of the form: const. times $x^{\alpha+\gamma-\zeta} f^{(\beta+\zeta)}(x)$, where $\zeta \le \delta$. Each of these is bounded, by definition, by one of our norms.

Theorem: The Fourier transformation is a continuous function on S.

Proof: We have to show that, for all (α,β) in $\mathbf{N}^{n2}$, there exists a finite subset $\Phi_{\alpha\beta}$ of $\mathbf{N}^{n2}$, and positive numbers $C^{\alpha\beta}_{\gamma\delta}$, such that for all f in S, $\| \hat{f} \|_{\alpha\beta} \le \sum_{(\gamma,\delta)\in\Phi} C^{\alpha\beta}_{\gamma\delta} \| f \|_{\alpha\beta}$. We have $\xi^\alpha \hat{f}^{(\beta)}(\xi) = \xi^\alpha((-ix)^\beta f(x))\,^\wedge(\xi) = (-i)^\alpha\left(D^\alpha((-ix)^\beta f(x))\right)^\wedge(\xi)$. It is enough to estimate the L^1-norm of $D^\alpha((-ix)^\beta f(x))$. To do this it is enough to estimate the supremum of $(1 + |x|^2)^n D^\alpha(x^\beta f(x))$. This last quantity is a linear combination of terms of the form $x^{2\gamma+\beta-\zeta} f^{(\zeta)}(x)$, where $\zeta \le \alpha$, $\zeta \le \beta$, and $|\gamma| \le n$. It is dominated by $\| f \|_{2\gamma+\beta-\zeta,\zeta}$. It follows that F is continuous on S.

the extension of linear operators to tempered distributions

The main point of interest in our study of tempered distributions is the extension of linear operators such as the Fourier transform and differentiation to tempered distributions. This involves a little work. The usual procedure is best illustrated with examples.

Example: Let us calculate the derivative, ***in the sense of distributions***, of [x > 0], the characteristic function of the positive real numbers. First, we express [x > 0] as a tempered distribution: $([x > 0], f) := \int [x > 0] f(x)\, dx$. Now if it were the case that [x > 0] were differentiable, the formula for integration by parts would give the equation

$$\int \frac{d}{dx} [x > 0] f(x)\, dx = -\int [x > 0] \frac{d}{dx} f(x)\, dx = f(0).$$

In this equation, only the first member does not make sense. Therefore, we make the members that do make sense the definition of the left side, expressed as a tempered distribution, rather than an integral:

$$(\frac{d}{dx} [x > 0], f) := -([x > 0], \frac{df}{dx}) = f(0).$$

Note that, when the tempered distribution pairs a suitably bounded function with a function in S, the meaning is the integral. In the present case, in terms of the delta function, we have $\frac{d}{dx} [x > 0] = \delta$. We can now differentiate the delta function: $(\frac{d}{dx} \delta, f) := -(\delta, \frac{df}{dx}) = -f'(0)$. We can think of δ' as a dipole… You can work out other examples. For example, what is the second derivative, in the sense of distributions, of $|x|$? As you experiment, you have to be careful sometimes. For example, the function $\log |x|$ is locally integrable, so $(\log |x|, f) := \int \log |x| f(x)\, dx$. Therefore, in the sense of distributions, $(\frac{d}{dx} \log |x|, f) := -(\log |x|, \frac{df}{dx})$. This is all very well; the fun comes in simplifying beyond the mere definition! We have (recalling improper Riemann integrals)

$$-(\log |x|, \frac{df}{dx}) = -\int \log |x| \frac{df}{dx}\, dx = -\lim_{b\uparrow 0,a\downarrow 0}\left(\int_{-\infty}^{b} \log |x| \frac{df}{dx} dx + \int_{a}^{\infty} \log |x| \frac{df}{dx} dx\right) =$$

$$= \lim_{b\uparrow 0,a\downarrow 0}\left(\int_{-\infty}^{b} \frac{1}{x} f(x)\, dx - \log |b|\, f(b) + \int_{a}^{\infty} \frac{1}{x} f(x)\, dx + \log |a|\, f(a)\right) = ?$$

The two integrals do not converge separately, at least for all f in S. The boundary terms don't converge either. The usual technique is to use the **Cauchy principal value**, the limit with $b = -a$, or

$$\lim_{\varepsilon\downarrow 0}\left(\int_{-\infty}^{-\varepsilon}\frac{1}{x}f(x)\,dx - \log|\varepsilon|\,f(-\varepsilon) + \int_{\varepsilon}^{\infty}\frac{1}{x}f(x)\,dx + \log|\varepsilon|\,f(\varepsilon)\right) =$$

$$= \lim_{\varepsilon\downarrow 0}\left(\int_{|x|>\varepsilon}\frac{1}{x}f(x)\,dx + \log|\varepsilon|\,(f(\varepsilon)-f(-\varepsilon))\right) =$$

$$= \lim_{\varepsilon\downarrow 0}\left(\int_{\varepsilon}^{\infty}\frac{f(x)-f(-x)}{x}\,dx + \log|\varepsilon|\,(f(\varepsilon)-f(-\varepsilon))\right) = \int_{0}^{\infty}\frac{f(x)-f(-x)}{x}\,dx.$$

The last integral is absolutely convergent because f has a continuous first derivative. The boundary terms tend to 0 by the mean value theorem and a basic limit: $x\log x \to 0$ as x decreases to 0. The Cauchy principal value in this case is usually written

$$\lim_{\varepsilon\downarrow 0}\int_{|x|>\varepsilon} f(x)\,\frac{dx}{x}\ .$$

Thus, in the sense of distributions, the derivative of $\log|x|$ is the tempered distribution

$$(\text{p.v.}\frac{1}{x}\ , f) := \lim_{\varepsilon\downarrow 0}\int_{|x|>\varepsilon} f(x)\,\frac{dx}{x}\ .$$

Exercise: Find the derivative, in the sense of distributions, of $\text{p.v.}\frac{1}{x}$. Your goal is to find an expression for it in terms of f, not in terms of f'. Suggestion: when you integrate by parts, subtract an appropriate constant of integration!

Example: Let us calculate the Fourier transform, ***in the sense of distributions***, of $[\,x > 0\,]$, the characteristic function of the positive real numbers. First, we express $[\,x > 0\,]$ as a tempered distribution: $([\,x > 0\,], f) := \int [\,x > 0\,]\, f(x)\,dx$. Now if it were the case that $[\,x > 0\,]$ were integrable, the equation $\int \hat{f}(\xi)\; g(\xi)\,d\xi = \int f(x)\,\hat{g}(x)\,dx$ would give $\int [\,x > 0\,]^\wedge(\xi)\, f(\xi)\,dx = \int [\,x > 0\,]\,\hat{f}(x)\,dx$. The right-hand side makes sense. We thus define the Fourier transform, in the sense of distributions, of $[\,x > 0\,]$, by the equation $([\,x > 0\,]^\wedge, f) := \int [\,x > 0\,]\,\hat{f}(x)\,dx$. Again, we want to express the integral on the right in terms of f, not $\hat{f}$. We can use the approximation device, and write $\int [\,x > 0\,]\,\hat{f}(x)\,dx = \lim_{\varepsilon\to 0}\int_0^{\infty} e^{-\varepsilon x}\,\hat{f}(x)\,dx$. Now we can interchange the order of integration.

We have $\int_0^{\infty} e^{-\varepsilon x}\,\hat{f}(x)\,dx = \int_0^{\infty} e^{-\varepsilon x}\int e^{-ixy}\, f(y)\,dy\,dx = \int \frac{f(y)}{\varepsilon + iy}\,dy$. The last integral resembles $(\text{p.v.}\frac{1}{x}\ , f)$. To exploit this, let us write

$$\int \frac{f(y)}{\varepsilon+iy}\,dy = \int_{|y|>\varepsilon} f(y)\left(\frac{1}{\varepsilon+iy}-\frac{1}{iy}\right)dy + \int_{|y|>\varepsilon}\frac{f(y)}{iy}\,dy + \int_{|y|\le\varepsilon}\frac{f(y)}{\varepsilon+iy}\,dy.$$

In the first and third integrals let us replace y by εy. This gives

$$\int \frac{f(y)}{\varepsilon+iy}\,dy = \int_{|y|>1} f(\varepsilon y)\left(\frac{1}{1+iy}-\frac{1}{iy}\right)dy + \int_{|y|>\varepsilon}\frac{f(y)}{iy}\,dy + \int_{|y|\le 1}\frac{f(\varepsilon y)}{1+iy}\,dy.$$

Now we can let $\varepsilon \to 0$, and use dominated convergence to get

$$\lim_{\varepsilon\to 0}\int_0^{\infty} e^{-\varepsilon x}\,\hat{f}(x)\,dx = \int_{|y|>1} f(0)\left(\frac{1}{1+iy}-\frac{1}{iy}\right) dy - i\,(\text{p.v.}\frac{1}{x}, f) + \int_{|y|\le 1}\frac{f(0)}{1+iy}\,dy.$$

It remains to evaluate the two integrals $\int_{|y|>1}\left(\frac{1}{1+iy}-\frac{1}{iy}\right) dy$ and $\int_{|y|\le 1}\frac{dy}{1+iy}$. Let us replace y by $-y$ in each of these integrals. We get $\int_{|y|>1}\left(\frac{1}{1-iy}+\frac{1}{iy}\right) dy$ and $\int_{|y|\le 1}\frac{dy}{1-iy}$. Now let us add corresponding integrals and divide by 2. We get $\int_{|y|>1}\frac{1}{1+y^2}\,dy$ and $\int_{|y|\le 1}\frac{dy}{1+y^2}$. The sum of these is π. Therefore, $([\,x>0\,]^\wedge, f) = (\pi\delta, f) - i\,(\text{p.v.}\frac{1}{x}, f)$. We say $[\,x>0\,]^\wedge = \pi\delta - i\,\text{p.v.}\frac{1}{x}$.

Exercise: Show that $\hat{\delta} = 1$; interpret the equation!

Definition: The partial derivative, $u^{(\alpha)}$, of a tempered distribution u, is defined to be

$$(u^{(\alpha)}, f) := (-1)^{|\alpha|}(u, f^{(\alpha)}).$$

Definition: The Fourier transform, $\hat{u}$, of a tempered distribution u, is defined to be

$$(\hat{u}, f) := (u, \hat{f}).$$

Each of these definitions uses the idea of the **transpose** of a linear transformation. This is defined, abstractly, as follows: Suppose that T is a linear transformation from a vector space V to itself. Let V^* denote a space of linear funcionals on V. Then, for each v^* in V^*, the equation $T^*v^*(v) := v^*(Tv)$ defines a linear functional on V. Moreover, the transpose, T^*, is a linear transformation. The only issue, in applications, is whether or not T^*v^* is in V^*. In our case, with $V = \mathcal{S}$, all we need is for T to be continuous from $\mathcal{S}$ to $\mathcal{S}$.

Exercise: Prove that each of these definitions yields a tempered distribution. Your proof must be correct and very short to be accepted!

There is another point to be emphasized here! We really need an operator whose transpose is known to us. Then *its* transpose, as a map from $\mathcal{S}'$ to $\mathcal{S}'$, will be the desired extension.

Exercise: State and prove the Fourier Inversion Theorem appropriate to $\mathcal{S}'$. You will need to define the “reflection” of a tempered distribution too. While at it, define the complex conjugate of a tempered distribution too.

Exercise: Find the Fourier transform of $D^\alpha\delta$.

Exercise: Find the Fourier transform of x^α.

Exercise: Show that the convolution of a tempered distribution and a function in $\mathcal{S}$ is a function in $\mathcal{M}$. The convolution is defined intuitively by $\int u(y)\,f(x-y)\,dy$, and *actually* by the equation

$$u*f(x) := (u, \tau_x Rf).$$

the topology of $\mathcal{S}$ is given by a metric.

There is a standard technique for combining a countable family of metrics d_k to make a topology. There are two steps. In first step we replace each metric by the equivalent, bounded, metric, $\frac{d_k(x, y)}{1 + d_k(x, y)}$. This is still a metric because the function $\frac{x}{1+x}$ is subadditive on $[0, \infty)$. Then we form the new metric $d(x, y) := \sum_k 2^{-k}\frac{d_k(x, y)}{1 + d_k(x, y)}$. It is a routine matter to show that d is a metric, and that the topology it determines is the smallest one that contains all the topologies determined by the individual metrics that determine d.

Here is a metric for $\mathcal{S}$: Let $d(f, g) := \sum_{(\alpha, \beta) \in \mathbf{N}^n \times \mathbf{N}^n} 2^{-|\alpha|-|\beta|-2n} \frac{\| f-g \|_{\alpha\beta}}{1 + \| f-g \|_{\alpha\beta}}$.

Theorem: A sequence $\{ f_k \}$ in $\mathcal{S}$ converges to g in $\mathcal{S}$ if and only if $\| f_k - g \|_{\alpha\beta} \to 0$ for each (α, β) in $\mathbf{N}^n \times \mathbf{N}^n$:

Proof: For a fixed (α, β) in $\mathbf{N}^n \times \mathbf{N}^n$,

$$d(f_k, g) \geq 2^{-|\alpha|-|\beta|-2n} \frac{\| f_k - g \|_{\alpha\beta}}{1 + \| f_k - g \|_{\alpha\beta}} \geq 2^{-|\alpha|-|\beta|-2n} \| f_k - g \|_{\alpha\beta};$$

on the other hand, Lebesgue’s dominated convergence theorem shows that the converse is true.

Example: Here is a continuous non-linear function from $(0, \infty)$ into $\mathcal{S}$, namely the approximate identity $f_\varepsilon(x) := f*\phi_\varepsilon(x)$, where $\phi_\varepsilon(x) := \varepsilon^{-n}\phi(x/\varepsilon)$, and $\phi(x)$ is a C^∞ function that is non-negative, even, has support in the unit ball, and whose integral is 1. We check most easily that f_ε belongs to $\mathcal{S}$ by taking the Fourier transform.

Theorem: f_ε converges to f in $\mathcal{S}$, as $\varepsilon \to 0$. $ApI(\varepsilon) := f_\varepsilon$ is continuous from $(0, \infty)$ into $\mathcal{S}$.

Proof: We can use the theorem just proved, and show two things: For each (α, β) in $\mathbf{N}^n \times \mathbf{N}^n$, $\| f_\varepsilon - f \|_{\alpha\beta} \to 0$, as $\varepsilon \to 0$, and $\| f_\varepsilon - f_\eta \|_{\alpha\beta} \to 0$, as $\varepsilon \to \eta$, for each $\eta > 0$.

The first of these follows a pattern that I hope is familiar by now. The idea is simple, but the details are tedious, and they make use of some facts from calculus that you might not be familiar with. Now,

$$\begin{aligned} x^\alpha f_\varepsilon^{(\beta)}(x) &= \int \phi_\varepsilon(y)\, (x - y + y)^\alpha f^{(\beta)}(x-y)\, dy \\ &= \int \phi_\varepsilon(y) \sum_{\gamma \leq \alpha} \binom{\alpha}{\gamma} (x-y)^\gamma y^{\alpha-\gamma} f^{(\beta)}(x-y)\, dy \\ &= \int \phi_\varepsilon(y) \sum_{\gamma \leq \alpha} \binom{\alpha}{\gamma} y^{\alpha-\gamma} \left((x-y)^\gamma f^{(\beta)}(x-y) - x^\gamma f^{(\beta)}(x) + x^\gamma f^{(\beta)}(x) \right) dy \end{aligned}$$

$$= \int \phi_\varepsilon(y) \sum_{\gamma \le \alpha} \binom{\alpha}{\gamma} y^{\alpha-\gamma}\big((x-y)^\gamma f^{(\beta)}(x-y) - x^\gamma f^{(\beta)}(x)\big)\ dy$$

$$+ \int \phi_\varepsilon(y) \sum_{\gamma \le \alpha} \binom{\alpha}{\gamma} y^{\alpha-\gamma} x^\gamma f^{(\beta)}(x)\ dy \ =:\ I + II, \text{ for short.}$$

Each of I and II is a function of x and ε. We need estimates for $x^\alpha f_\varepsilon^{(\beta)}(x) - x^\alpha f^{(\beta)}(x)$.

We have $II - x^\alpha f^{(\beta)}(x) = \int \phi_\varepsilon(y) \sum_{\gamma \le \alpha,\ \gamma \ne \alpha} \binom{\alpha}{\gamma} y^{\alpha-\gamma} x^\gamma f^{(\beta)}(x)\, dy$. For each term in the sum, we have

$$\left| \int \phi_\varepsilon(y)\binom{\alpha}{\gamma} y^{\alpha-\gamma} x^\gamma f^{(\beta)}(x)\, dy \right| \le \int \phi_\varepsilon(y)\binom{\alpha}{\gamma} |y^{\alpha-\gamma}|\ \|f\|_{\gamma\beta}\, dy.$$

Recall that $|y^{\alpha-\gamma}| \le |y|^{|\alpha|-|\gamma|}$. Since $\gamma \le \alpha$ and $\gamma \ne \alpha$, the exponent $|\alpha| - |\gamma| \ge 1$.

Since the $|y| \le \varepsilon$ are the only ones that matter, we can write $|y^{\alpha-\gamma}| \le \varepsilon$.

Thus,

$$\left| \int \phi_\varepsilon(y)\binom{\alpha}{\gamma} y^{\alpha-\gamma} x^\gamma f^{(\beta)}(x)\, dy \right| \le \binom{\alpha}{\gamma} \varepsilon\ \|f\|_{\gamma\beta} \int \phi_\varepsilon(y)\, dy = \binom{\alpha}{\gamma} \varepsilon\ \|f\|_{\gamma\beta}.$$

To handle the term I, let us look at the sum in the integrand one term at a time. In each term, we want to estimate the factor $(x-y)^\gamma f^{(\beta)}(x-y) - x^\gamma f^{(\beta)}(x)$. This is a complex-valued function. We will use the Mean Value Inequality on this difference, viewed as a function of y.

The Mean Value Theorem is not true for complex-valued functions of a real variable. Here is an example. Let $F(t) = e^{i\pi t}$, $0 \le t \le 1$. Then $F'(t) = i\pi\, e^{i\pi t}$, but $F(1) - F(0) = -2 \ne F'(t)(1-0)$ for any value of t, since $|F'(t)| = \pi \ne 2$, for all t. Nevertheless, we have, for any continuously differentiable (in the real sense) complex-valued function F of $n \ge 1$ variables, the estimate, known as the **Mean Value Inequality**: $|F(y) - F(x)| \le \sup_{Sxy} |\text{grad } F|\ |y - x|$, where Sxy is the line segment that joins x and y, $|\text{grad } F| := \left(\sum_{j=1}^{n} \left| \frac{\partial F}{\partial x_j} \right|^2 \right)^{1/2}$, and each partial derivative is evaluated at a point on Sxy. We prove this by forming the function $g(t) := F(x + t(y-x))$, $0 \le t \le 1$, which has derivative $g'(t) = \text{grad } F(x + t(y-x)) \cdot (y - x) := \sum_{j=1}^{n} \frac{\partial F}{\partial x_j}(y_j - x_j)$, where the partial derivatives are evaluated at $x + t(y-x)$. Then we apply the Fundamental Theorem of Calculus, Schwarz' Inequality, and estimate $|\text{grad } F|$ by the supremum that appears in the desired inequality. If it is convenient, we can use $\sum_{j=1}^{n} \left| \frac{\partial F}{\partial x_j} \right|$ instead of the quadratic norm, by the triangle inequality.

Put $G(y) := (x-y)^\gamma f^{(\beta)}(x-y)$. We will estimate $G(y) - G(0)$. At y, by the Chain Rule, $\frac{\partial G}{\partial y_j} = -\gamma_j (x-y)^{\gamma - e_j} f^{(\beta)}(x-y) - (x-y)^\gamma f^{(\beta+e_j)}(x-y)$; the first term is absent if $\gamma_j = 0$. We have $\left| \frac{\partial G}{\partial y_j} \right| \le \gamma_j \|f\|_{\gamma - e_j, \beta} + \|f\|_{\gamma, \beta + e_j}$. Since the $|y| \le \varepsilon$ are the only ones that matter,

$\left|\frac{\partial G}{\partial y_j}\right| \le \gamma_j \,\|f\|_{\gamma-e_j,\beta} + \|f\|_{\gamma,\beta+e_j}$. We can skip some details here. When these estimates are added, and multiplied by $|y|$, since the $|y| \le \varepsilon$ are the only ones that matter, we see that term I is dominated by a large constant, times terms involving norms $\|f\|_{\gamma\delta}$, times ε. We now combine our estimates, and conclude that $\|f_\varepsilon - f\|_{\alpha\beta} \to 0$, as $\varepsilon \to 0$. This is so for each α and β, so by the theorem, $f_\varepsilon - f \to 0$, as $\varepsilon \to 0$, as desired.

The second task is to show $\|f_\varepsilon - f_\eta\|_{\alpha\beta} \to 0$, as $\varepsilon \to \eta$, for each $\eta > 0$. We have

$$x^\alpha f_\varepsilon^{(\beta)}(x) - x^\alpha f_\eta^{(\beta)}(x) = \int (\phi_\varepsilon(y) - \phi_\eta(y))\,(x - y + y)^\alpha f^{(\beta)}(x-y)\,dy$$
$$= \int (\phi_\varepsilon(y) - \phi_\eta(y)) \sum_{\gamma \le \alpha} \binom{\alpha}{\gamma} y^{\alpha-\gamma}(x-y)^\gamma f^{(\beta)}(x-y)\,dy.$$

Again, let us estimate the sum term by term. We have $\left|\int (\phi_\varepsilon(y) - \phi_\eta(y))\, y^{\alpha-\gamma}(x-y)^\gamma f^{(\beta)}(x-y)\,dy\right| \le \int |\phi_\varepsilon(y) - \phi_\eta(y)|\,|y^{\alpha-\gamma}|\,\|f\|_{\gamma\beta}\,dy$. It is therefore enough to show that $\int \left|\frac{1}{\varepsilon^n}\phi\left(\frac{y}{\varepsilon}\right) - \frac{1}{\eta^n}\phi\left(\frac{y}{\eta}\right)\right| |y^{\alpha-\gamma}|\,dy$ tends to 0 as ε tends to η, for each positive η. Upon multiplying y by ε, we have

$\varepsilon^{|\alpha|-|\gamma|}\int \left|\phi(y) - \frac{\varepsilon^n}{\eta^n}\phi\left(\frac{\varepsilon y}{\eta}\right)\right| |y^{\alpha-\gamma}|\,dy$. Since ε is close to η, let's subtract and add $\phi\left(\frac{\varepsilon y}{\eta}\right)$. This gives, after an application of the triangle inequality,

$\varepsilon^{|\alpha|-|\gamma|}\int \left|\phi(y) - \phi\left(\frac{\varepsilon y}{\eta}\right)\right| |y^{\alpha-\gamma}|\,dy + \varepsilon^{|\alpha|-|\gamma|}\int \left|(1 - \frac{\varepsilon^n}{\eta^n})\,\phi\left(\frac{\varepsilon y}{\eta}\right)\right| |y^{\alpha-\gamma}|\,dy =: I + II$. In the second integral, let's multiply y by $\frac{\eta}{\varepsilon}$. This gives $II = \left(\frac{\eta}{\varepsilon}\right)^n \eta^{|\alpha|-|\gamma|}\int \left|(1 - \frac{\varepsilon^n}{\eta^n})\,\phi(y)\right| |y^{\alpha-\gamma}|\,dy$. As ε approaches η, II approaches 0, uniformly in x.

Finally, we turn to term I. Let us estimate $\left|\phi(y) - \phi\left(\frac{\varepsilon y}{\eta}\right)\right|$. Consider the function of t defined, for t close to 1, by $g(t) := \phi(ty)$. The Mean Value Theorem can be applied; we get $\phi(y) - \phi(ty) = (\text{grad}\,\phi(\tau y))\cdot y\,(1 - t)$, for some τ in $(t, 1)$. Since ϕ is in $\mathcal{S}$, there exists a constant C such that $|\,\text{grad}\,\phi(\tau y)\,| \le \frac{C}{(1+|y|^2)^{n+1+|\alpha|}}$, if $|\tau - 1| \le 1/2$. This all gives the estimate

$$I \le \varepsilon^{|\alpha|-|\gamma|}\int \frac{C|y|(1 - \frac{\varepsilon}{\eta})}{(1+|y|^2)^{n+1+|\alpha|}}\,|y^{\alpha-\gamma}|\,dy, \text{ which tends to } 0 \text{ as } \varepsilon \to \eta.$$

This was enough; the theorem is proved.

a remark on the theorem

The restriction to approximate identities based on a function with support in the unit ball is not necessary, only convenient. Without it, there is always a second term that takes advantage of the smallness of the integral away from the origin.

the weak*-topology of $\mathcal{S}'$.

There is a standard topology for dual spaces, called the **weak*-topology**. Sometimes it is the only one used in practice. This topology consists of arbitrary unions of translates of finite intersections of sets of the form $\{ v^* \in V^* : |v^*(v)| < \varepsilon \}$, where v is an arbitrary vector in V, and ε is a positive number. Thus, a basic neighborhood of 0 in V^* is a set of the form

$$\{ v^* \in V^* : |v^*(v_k)| < \varepsilon_k, \text{ for } k = 1, \dots , N < \infty \}.$$

This is the smallest topology such that all the evaluation maps $v^* \to v^*(v)$, v in V, are continuous on V^*. With respect to the weak*-topology, the transpose of a continuous linear map on $\mathcal{S}$ is continuous. The proof of this is a matter of checking definitions.

A fact that is more specific to $\mathcal{S}$ is that $\mathcal{M}$, viewed as a subset of $\mathcal{S}'$ by the map that associates m in $\mathcal{M}$ with the linear functional $(m, g) := \int m(x)\, g(x)\, dx$, is weak*-dense in $\mathcal{S}'$. As you work with $\mathcal{S}$ and $\mathcal{S}'$, this may help dispel the uneasiness that comes with enjoying "magic." To show this density we can use approximate identities, and put the convolution on the distribution. To be sure this will work, we need to check that the convolution of a distribution and a function in $\mathcal{S}$ is in $\mathcal{M}$. We need to define the convolution $u * f$, u in $\mathcal{S}'$, f in $\mathcal{S}$.

Definition: The **convolution**, $u * f$, of u in $\mathcal{S}'$, f in $\mathcal{S}$ is the function $u * f(x) := (u, f(x - \bullet))$, that is, u applied to the function $f(x - y)$, as a function of y; x is regarded as a constant. We can also write $u * f$ as $u * f(x) = (u, \tau_x Rf)$.

Theorem: If u is in $\mathcal{S}'$, and f is in $\mathcal{S}$, then $u * f$ is in $\mathcal{M}$.

Proof: We know that there exists a finite subset Φ of $\mathbf{N}^{n2}$, and positive numbers $C_{\alpha\beta}$, such that for all f in $\mathcal{S}$,

$$|(u, f)| \le \sum_{(\alpha,\beta)\in\Phi} C_{\alpha\beta} \, \| f \|_{\alpha\beta}.$$

Therefore, $|u * f(x)| \le \sum_{(\alpha,\beta)\in\Phi} C_{\alpha\beta} \| \tau_x Rf \|_{\alpha\beta}$. We need to estimate $\| \tau_x Rf \|_{\alpha\beta}$. Now,

$\| \tau_x Rf \|_{\alpha\beta} = \sup_y |y^\alpha f^{(\beta)}(x - y)|$. As usual by now, we write

$$y^\alpha f^{(\beta)}(x - y) = \sum_{\gamma \le \alpha} \binom{\alpha}{\gamma} x^{\alpha-\gamma} (y - x)^\gamma f^{(\beta)}(x - y).$$

Let us define $w := (|x_1|, \dots , |x_n|)$, $\mathbf{1} := (1, \dots , 1)$. Then we have

$$|y^\alpha f^{(\beta)}(x - y)| \le \sum_{\gamma \le \alpha} \binom{\alpha}{\gamma} w^{\alpha-\gamma} \| f \|_{\gamma\beta} \le \max_{\gamma \le \alpha} \| f \|_{\gamma\beta} \, (w + \mathbf{1})^\alpha \le \max_{\gamma \le \alpha} \| f \|_{\gamma\beta} \, 2^{|\alpha|}(1 + |x|^2)^{|\alpha|}.$$

It follows that $u * f$ is bounded by a polynomial.

> Note: Here, as elsewhere in these notes, I have avoided the refinement that the exponent $|\alpha|$ in the last expression could be replaced by $[|\alpha|/2] + 1$, or even by "the smallest integer whose double is at least $|\alpha|$."

It remains to show that $u * f$ is C^∞, and that each derivative is bounded by a polynomial. If we can show, for each k, that $\dfrac{\partial(u*f)}{\partial x_k}$ exists, and is equal to $u*\dfrac{\partial f}{\partial x_k}$, we will be done because of what has just been proved.

Exercise: Prove that $\dfrac{\partial(u*f)}{\partial x_k}$ exists, and is equal to $u*\dfrac{\partial f}{\partial x_k}$. The only way I know to do this is to show that the difference quotient for the partial derivative converges in S to the partial derivative.

Theorem: $\mathcal{M}$ is dense in $\mathcal{S}'$.

> What this means is that the techniques we used in the examples amount to working in a dense set, where we show continuity, and then extend to the closure, "by continuity."

Proof: We seek to show that there is an $\mathcal{M}$ - functional m^* in every set of the form

$$\{ u \in \mathcal{S}' : |u(f_k) - u_o(f_k)| < \varepsilon_k, \text{ for } k = 1, \dots, N < \infty \}.$$

To do this we use an approximate identity. Set $u_\varepsilon(x) := u_o*\phi_\varepsilon(x)$. Then u_ε is in $\mathcal{M}$, and we consider $\int u_\varepsilon(x)\, f_k(x)\, dx$ as the limit of Riemann sums over larger and larger cubes with finer and finer meshes. Each such sum can be written

$$\sum_p u_\varepsilon(x_p)\, f_k(x_p)\, \lambda(Q_p) = (u_o, \sum_p \phi_\varepsilon(x_p - \bullet)\, f_k(x_p)\, \lambda(Q_p))$$
$$= (u_o, \sum_p \phi_\varepsilon(\bullet - x_p)\, f_k(x_p)\, \lambda(Q_p)),$$

because ϕ is even.

Exercise: Prove that $\sum_p \phi_\varepsilon(y - x_p)\, f_k(x_p)\, \lambda(Q_p)$ converges in $\mathcal{S}$ to $\phi_\varepsilon * f_k$. You also need to specify the sizes of the cubes and their meshes. Dyadic subdivision is the easiest to deal with. It is also convenient to express the sum as the integral of a function that is constant on small subcubes.

Then we have $\int u_\varepsilon(x)\, f_k(x)\, dx = (u_o, \phi_\varepsilon * f_k) \to (u_o, f_k)$. Now we can choose ε so small that each of the inequalities $|u_\varepsilon(f_k) - u_o(f_k)| < \varepsilon_k$ holds. Let $m^* = u_\varepsilon$. This establishes the density.