MATH 172: TEMPERED DISTRIBUTIONS AND THE FOURIER TRANSFORM

ANDRAS VASY

We have seen that the Fourier transform is well-behaved in the framework of Schwartz functions as well as L^2, while L^1 is much more awkward. Tempered distributions, which include L^1, provide a larger framework in which the Fourier transform is well-behaved, and they provide the additional benefit that one can differentiate them arbitrarily many times!

To see how this is built up, we start with a reasonable class of objects, such as bounded continuous functions on $\mathbb{R}^n$, and embed them into a bigger space by a map ι. The bigger space is that of tempered distributions, which we soon define. The idea is that ι is a one-to-one map, thus we may think of bounded continuous functions as tempered distributions by identifying $f \in C^0_\infty(\mathbb{R}^n)$ with $\iota(f)$. (Below we write $\iota(f) = \iota_f$ often, by analogy of the notation of a sequence, as a distribution itself will be a map, or function, on functions, so we need to write expressions like $\iota_f(\phi)$, which is nicer than $(\iota(f))(\phi)$.) An analogy is that letters of the English alphabet can be considered as numbers via their ASCII encoding; there are more ASCII codes than letters, but to every letter corresponds a unique ASCII code. One can just think of letters then as numbers, e.g. one thinks of the letter 'A' as the decimal number 65, i.e. the letters are thought of as a subset of the integers 0 through 255.

So on to distributions. Suppose V is a vector space over $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$. The algebraic dual of V is the vector space $\mathcal{L}(V, \mathbb{F})$ consisting of linear functionals from V to $\mathbb{F}$. That is elements of $f \in \mathcal{L}(V, \mathbb{F})$ are linear maps $f : V \to \mathbb{F}$ satisfying

$$f(v + w) = f(v) + f(w),\ f(cv) = cf(v),\ v, w \in V,\ c \in \mathbb{F}.$$

When V is infinite dimensional, we need additional information, namely continuity. So if V is a topological space (which means that one has a notions of *open* sets, which we do not emphasize, and thus of *convergence*, which we do), indeed in our case a metric space, with the topology compatible with the vector space structure (namely the vector space operations are continuous, which we have already shown in the case of normed vector spaces), i.e. if V is a topological vector space, we define the dual space V^* as the space of *continuous* linear maps $f : V \to \mathbb{F}$; in the cases we are interested in, continuity is the same as *sequential continuity.* (This is always the case in metric spaces!) The latter means that we consider maps f with the property that if $\phi_j \to \phi$ in V then $f(\phi_j) \to f(\phi)$ in $\mathbb{F}$.

For us, V is the class of 'very nice objects', and V^* will be the class of 'bad objects'. Of course, normally there is no way of comparing elements of V with those of V^*, so we will also need an injection (i.e. a one-to-one map)

$$\iota : V \to V^*$$

so that elements of V can be regarded as elements of V^* (by identifying $v \in V$ with $\iota(V)$). As we want to differentiate functions, as much as we desire, V will consist of infinitely differentiable functions. As we need to control behavior at infinity to integrate, the elements of V must decay at infinity. Concretely, as we want good behavior relative to the Fourier transform, we take V to be the set of Schwartz functions.

One may also call elements of V *test functions* since V^* is defined as continuous linear maps from V to $\mathbb{F}$, i.e. we are applying elements of V^* to elements of V, so we are 'testing' elements of V^* on V. This is more common, however, when V is taken to be the set of compactly supported $\mathcal{C}^\infty$ functions, $\mathcal{C}^\infty_c(\mathbb{R}^n)$; in that case the dual objects in V^* are called *distributions*.

In order to motivate the definition of $\mathcal{S}$-convergence, recall that $\mathcal{S} = \mathcal{S}(\mathbb{R}^n)$ is the set of functions $\phi \in \mathcal{C}^\infty(\mathbb{R}^n)$ with the property that for any multiindices $\alpha, \beta \in \mathbb{N}^n$, $x^\alpha \partial^\beta \phi$ is bounded. Here we wrote $x^\alpha = x_1^{\alpha_1} x_2^{\alpha_2} \dots x_n^{\alpha_n}$, and $\partial^\beta = \partial_{x_1}^{\beta_1} \dots \partial_{x_n}^{\beta_n}$; with $\partial_{x_j} = \frac{\partial}{\partial x_j}$.

With this in mind, convergence of a sequence $\phi_m \in \mathcal{S}$, $m \in \mathbb{N}$, to some $\phi \in \mathcal{S}$, in $\mathcal{S}$ is defined as follows. We say that ϕ_m converges to ϕ in $\mathcal{S}$ if for all multiindices α, β, $\sup |x^\alpha \partial^\beta (\phi_m - \phi)| \to 0$ as $m \to \infty$, i.e. if $x^\alpha \partial^\beta \phi_m$ converges to $x^\alpha \partial^\beta \phi$ uniformly.

This notion in fact arises from a metric. Let

$$\|\phi\|_N = \sup_{|\alpha|+|\beta| \leq N} \sup_{\mathbb{R}^n} |x^\alpha \partial^\beta \phi|.$$

Then for all N, ϕ_N is a norm on $\mathcal{S}(\mathbb{R}^n)$, with respect to which, however, $\mathcal{S}$ is not complete. (E.g. each norm controls only finitely many derivatives.) Let d_N be the corresponding metric:

$$d_N(\phi, \psi) = \|\phi - \psi\|_N, \ \phi, \psi \in \mathcal{S}(\mathbb{R}^n).$$

For us it is now convenient to 'cut d_N down to size', namely let

$$\tilde{d}_N(\phi, \psi) = \min(1, d_N(\phi, \psi)).$$

Note that $\tilde{d}_N$ is still a metric, and it is equivalent to d_N (has the same open sets, or equivalently, the same convergent sequences). Finally let

$$d(\phi, \psi) = \sum_{N=0}^{\infty} 2^{-N} \tilde{d}_N(\phi, \psi).$$

Note that the sum converges since $\tilde{d}_N \leq 1$ for all N. Then it is straightforward to check that d is a metric, using that $\tilde{d}_N$ is for each N, and further that convergence with respect to d is the same as the convergence *for every* $\tilde{d}_N$*, i.e. for every* d_N. Here the point is that:

- for any N and $\epsilon > 0$ there is $\epsilon' > 0$ such that $d(\phi, \psi) < \epsilon'$ implies $\tilde{d}_N(\phi, \psi) < \epsilon$ (indeed, take $\epsilon' = 2^{-N}\epsilon$), and
- conversely, for any $\epsilon' > 0$ there exists N and $\epsilon > 0$ such that $\tilde{d}_N(\phi, \psi) < \epsilon$ implies $d(\phi, \psi) < \epsilon'$ (this uses $d_m \leq d_N$ if $m \leq N$, and similarly for $\tilde{d}$; one then simply picks N so that the tail of the series, beyond N, is $< \epsilon'/2$, and then picks $\epsilon > 0$ so that the first N terms being $< \epsilon 2^{-N}$, respectively, sum up to $< \epsilon'/2$).

We can now define tempered distributions:

Definition 1. A *tempered distribution* $u \in \mathcal{S}'(\mathbb{R}^n)$ is a continuous linear functional on $\mathcal{S}(\mathbb{R}^n)$. That is, a tempered distribution u is a map $u : \mathcal{S}(\mathbb{R}^n) \to \mathbb{F}$ such that

(i) u is linear: $u(\phi + \psi) = u(\phi) + u(\psi)$, $u(c\phi) = cu(\phi)$ for $\phi, \psi \in \mathcal{S}(\mathbb{R}^n)$ and $c \in \mathbb{F}$,
(ii) and u is continuous, so if ϕ_j is any sequence such that $\phi_j \to \phi$ in $\mathcal{S}(\mathbb{R}^n)$ then $u(\phi_j) \to u(\phi)$ in $\mathbb{F}$.

It is straightforward to check (by linearity) that continuity of u is equivalent to the following: there exist N and $C > 0$ such that for all $\phi \in \mathcal{S}(\mathbb{R}^n)$,

$$|u(\phi)| \leq C\|\phi\|_N = C \sum_{|\alpha|+|\beta|\leq N} \sup_{\mathbb{R}^n} |x^\alpha \partial^\beta \phi|. \tag{1}$$

The simplest tempered distribution is the delta-distribution at some point $a \in \mathbb{R}^n$; it is given by

$$\delta_a(\phi) = \phi(a).$$

One can also generate many similar examples, e.g. the map $u : \mathcal{S}(\mathbb{R}^n) \to \mathbb{C}$ given by $u(\phi) = (\partial_1\phi)(a) - (\partial_2^2\phi)(b)$, where $a, b \in \mathbb{R}^n$, $n \geq 2$, is also a tempered distribution.

A large class of distributions is obtained the following way. Suppose that $f \in L^1(\mathbb{R}^n)$. Then f defines a distribution $u = \iota_f$ as follows:

$$u(\phi) = \iota_f(\phi) = \int_{\mathbb{R}^n} f\phi\,dx.$$

Note that the integral converges since $\phi \in \mathcal{S}(\mathbb{R}^n)$, and is thus bounded. Certainly $\iota_f : \mathcal{S}(\mathbb{R}^n) \to \mathbb{C}$ is linear; its continuity can be seen as follows. Suppose that $\phi_j \to \phi$ in $\mathcal{S}(\mathbb{R}^n)$, and thus in particular uniformly. Thus,

$$\begin{aligned}\Big|\int_{\mathbb{R}^n} f(x)\phi_j(x)\,dx - \int_{\mathbb{R}^n} f(x)\phi(x)\,dx\Big| &= \Big|\int_{\mathbb{R}^n} f(x)(\phi_j(x) - \phi(x))\,dx\Big| \\ &\leq \int_{\mathbb{R}^n} |f(x)|\,|\phi_j(x) - \phi(x)|\,dx \leq \|f\|_{L^1} \sup|\phi_j - \phi|,\end{aligned}$$

so the uniform convergence of ϕ_j to ϕ gives the desired conclusion. (The equivalent characterization of continuity, in terms of (1) means that we could have simply checked $|\int f\phi| \leq C\|\phi\|_0$, which holds with $C = \|f\|_{L^1}$.)

The important fact is that we do not lose any information by thinking of elements of L^1 as distributions, i.e. the map ι is one-to-one.

Lemma 0.1. *The map $\iota : L^1(\mathbb{R}^n) \to \mathcal{S}'(\mathbb{R}^n)$ is injective.*

Because of this lemma, we can consider $L^1(\mathbb{R}^n)$ as a subset of $\mathcal{S}'(\mathbb{R}^n)$, via the identification ι.

Proof. Indeed, if $\iota_f = 0$ then $\int f\phi\,dx = 0$ for all $\phi \in \mathcal{S}(\mathbb{R}^n)$, which is easily seen to imply $f = 0$. One way to do this is to let $\phi_{t,y}(x) = K_t(y - x)$ be given by the heat kernel K_t for $t > 0$. We have already seen that $K_t * f \to f$ in L^1. But $(K_t * f)(y) = \iota_f(\phi_{t,y})$, so if $\iota_f = 0$, then $K_t * f = 0$ for all $t > 0$, and thus taking the limit, $t \to 0$, $f = 0$. (We could of course use any other family of good kernels whose elements are Schwartz.) $\square$

Note that restricted to $C(\mathbb{R}^n)$ with a polynomial bound the injectivity argument does not involve anything remotely sophisticated: if $f \in C(\mathbb{R}^n)$, and $f(x_0) \neq 0$ for some $x_0 \in \mathbb{R}^n$, by the continuity of f, for $\epsilon > 0$ sufficiently small, $|f(x) - f(x_0)| < |f(x_0)|/2$ for $|x - x_0| < \epsilon$. Now let ϕ be as in Lemma 0.1 of the basic Fourier transform notes, so $|f(x) - f(x_0)| < |f(x_0)|/2$ on $\operatorname{supp}\phi$. So

$$\Big|\int f\phi\,dx - f(x_0)\int \phi\,dx\Big| \leq \int |f(x) - f(x_0)|\,\phi(x)\,dx \leq \frac{|f(x_0)|}{2}\int \phi\,dx,$$

so

$$\Big|\int f\phi\,dx\Big| \geq \Big|f(x_0)\int \phi(x)\,dx\Big| - \Big|\int f\phi\,dx - f(x_0)\int \phi\,dx\Big| \geq \frac{|f(x_0)|}{2}\int \phi\,dx,$$

so $\int f\phi\,dx \neq 0$ as $\int \phi\,dx > 0$.

More generally, even functions f for which $(1+|x|^2)^{-N} f \in L^1(\mathbb{R}^n)$ are tempered distributions since for $\phi \in \mathcal{S}(\mathbb{R}^n)$, $(1+|x|^2)^N \phi$ is bounded, thus

$$\int_{\mathbb{R}^n} f\phi = \int_{\mathbb{R}^n} ((1+|x|^2)^{-N} f)((1+|x|^2)^N \phi)$$

can be analyzed as above. In particular any continuous function f satisfying an estimate $|f(x)| \leq C(1+|x|)^N$ for some N and C defines a tempered distribution $u = \iota_f$ via

$$u(\psi) = \iota_f(\psi) = \int_{\mathbb{R}^n} f(x)\psi(x)\,dx, \quad \psi \in \mathcal{S},$$

since

$$\Big|\int_{\mathbb{R}^n} f(x)\psi(x)\,dx\Big| \leq CM \int_{\mathbb{R}^n} (1+|x|)^N (1+|x|)^{-N-n-1}\,dx < \infty,$$
$$M = \sup\big((1+|x|)^{N+n+1}|\psi|\big) < \infty.$$

This is the reason for the 'tempered' terminology: the growth of f is 'tempered' at infinity.

A more extreme example, related to Poisson summation, is the following, on $\mathbb{R}$ for simplicity: Let $a > 0$, and let

$$u = \sum_{k\in\mathbb{Z}} \delta_{ak},$$

i.e. for $\phi \in \mathcal{S}(\mathbb{R})$, let

$$u(\phi) = \sum_{k\in\mathbb{Z}} \phi(ak).$$

This sum converges, since $|\phi(x)| \leq C(1+|x|^2)^{-1}$ as $\phi \in \mathcal{S}$, and $\sum_{k\in\mathbb{Z}} C(1+a^2k^2)^{-1}$ converges. It is also continuous since if $\phi_j \to \phi$ in $\mathcal{S}$, then $(1+x^2)\phi_j \to (1+x^2)\phi$ uniformly on $\mathbb{R}$, hence

$$\begin{aligned}|u(\phi_j) - u(\phi)| &\leq \sum_{k\in\mathbb{Z}} (1+a^2k^2)|\phi_j(ak) - \phi(ax)|(1+a^2k^2)^{-1} \\ &\leq \sum_{k\in\mathbb{Z}} \sup_{x\in\mathbb{R}} \big((1+x^2)|\phi_j(x) - \phi(x)|\big)\,(1+a^2k^2)^{-1} \\ &\leq \sup_{x\in\mathbb{R}} \big((1+x^2)|\phi_j(x) - \phi(x)|\big) \sum_{k\in\mathbb{Z}} (1+a^2k^2)^{-1},\end{aligned}$$

and the last sum converges, so $|u(\phi_j) - u(\phi)| \to 0$ as $j \to \infty$.

One usually equips $\mathcal{S}'(\mathbb{R}^n)$ with the so-called weak-* topology:

Definition 2. One says that a sequence $u_j \in \mathcal{S}'(\mathbb{R}^n)$ converges to $u \in \mathcal{S}'(\mathbb{R}^n)$ if $u(\phi) = \lim_{j\to\infty} u_j(\phi)$ for all $\phi \in \mathcal{S}(\mathbb{R}^n)$.

A word of warning: this is *not* the topology arising from a metric space, although *there is a topology* in which this is the notion of convergence.

As an example, fix $\epsilon = 1$ and let ϕ be as in Lemma 0.1 of the basic Fourier transform notes with $x_0 = 0$. Let δ_j be a sequence of positive constants with $\lim_{j\to\infty} \delta_j = 0$. Let u_j be the distribution given by $\delta_j^{-n}\phi(./\delta_j)$, i.e.

$$u_j(\psi) = \int \delta_j^{-n}\phi(x/\delta_j)\psi(x)\,dx.$$

Let $c = \int \phi\,dx$. Then $\lim_{j\to\infty} u_j = c\delta_0$. Indeed, for $c = 1$ this follows from $\phi_\delta = \delta^{-n}\phi(./\delta)$ forming a family of good kernels in the sense we discussed earlier.

To be more explicit, note that by the continuity of ψ, given $\epsilon' > 0$, there is $\delta' > 0$ such that $|x| < \delta'$ implies $|\psi(x) - \psi(x_0)| < \epsilon'$. Then

$$\begin{aligned}|u_j(\psi) - c\delta_0(\psi)| &= \Big| \int \delta_j^{-n}\phi(x/\delta_j)\psi(x)\,dx - \psi(0)\int \phi(x)\,dx\Big| \\ &= \Big| \int \delta_j^{-n}\phi(x/\delta_j)\psi(x)\,dx - \int \delta_j^{-n}\phi(x/\delta_j)\psi(0)\,dx\Big| \\ &\leq \int \delta_j^{-n}\phi(x/\delta_j)|\psi(x) - \psi(x_0)|\,dx.\end{aligned}$$

For j sufficiently large, $\delta_j < \delta'$, and $\phi(x/\delta_j) = 0$ if $|x|/\delta_j \geq 1$, i.e. if $x \geq \delta_j$, so certainly if $x \geq \delta'$. Correspondingly, in the integral, one can restrict the integration to $|x| \leq \delta_j$, where $|\psi(x) - \psi(x_0)| < \epsilon'$ to deduce that

$$|u_j(\psi) - c\delta_0(\psi)| \leq \epsilon' \int \delta_j^{-n}\phi(x/\delta_j)\,dx = c\epsilon'$$

for j sufficiently large. This proves our claim.

This is a rather typical example, and it is not hard to show that one can approximate any $u \in \mathcal{S}'(\mathbb{R}^n)$ in the weak-* topology by u_j which are given by $\mathcal{S}(\mathbb{R}^n)$ functions, i.e. $\mathcal{S}(\mathbb{R}^n)$ is *sequentially dense* in $\mathcal{S}'(\mathbb{R}^n)$ (this is actually a much stronger statement than being dense since $\mathcal{S}'$ is not a metric space).

We can now consider differentiation. The idea is that we already know what the derivative of a C^1 function is, so we should express it as a distribution in such a way that it obviously extends to the class of all distributions. Now, for $f \in C^1(\mathbb{R}^n)$ with $|f(x)|, |(\partial_j f)(x)| \leq C(1+|x|)^N$ for some N, the distribution associated to the function $\partial_j f$ satisfies

$$\iota_{\partial_j f}(\phi) = \int \partial_j f\,\phi\,dx = -\int f\,\partial_j\phi\,dx = -\iota_f(\partial_j\phi)$$

for all $\phi \in \mathcal{S}(\mathbb{R}^n)$. Motivated by this, we make the definition:

Definition 3. The partial derivatives of $u \in \mathcal{S}'(\mathbb{R}^n)$ are defined by

$$\partial_j u(\phi) = -u(\partial_j\phi).$$

Note that for $\phi \in \mathcal{S}(\mathbb{R}^n)$, $\partial_j\phi \in \mathcal{S}(\mathbb{R}^n)$, so this definition makes sense.

It is straightforward to check that $\partial_j u$ *is* a distribution, in particular is continuous as a map $\mathcal{S}(\mathbb{R}^n) \to \mathbb{C}$. Note also that this is the only reasonable notion of a derivative as the map $u \mapsto \partial_j u$ is continuous, i.e. $u_k \to u$ implies $\partial_j u_k \to \partial_j u$, and $\mathcal{S}(\mathbb{R}^n)$, on which we already know ∂_j, is dense in $\mathcal{S}'(\mathbb{R}^n)$.

Some examples: on $\mathbb{R}$,

$$\delta_a'(\phi) = -\delta_a(\phi') = -\phi'(a),\ \phi \in \mathcal{S}(\mathbb{R}).$$

Also, if H is the Heaviside step function, so $H(x) = 1$ for $x \geq 0$, $H(x) = 0$ for $x < 0$, then

$$H'(\phi) = (\iota_H)'(\phi) = -\iota_H(\phi') = -\int_0^\infty \phi'(x)\,dx = \phi(0)$$

for all $\phi \in \mathcal{S}(\mathbb{R})$, where we used the fundamental theorem of calculus. Therefore $H' = \delta_0$.

Note also that tempered distributions u may be multiplied by $\mathcal{C}^\infty$ functions g all of whose derivatives are polynomially bounded. Indeed, we proceed again by analogy with ι_f where $f \in C(\mathbb{R}^n)$. In that case

$$\iota_{fg}(\phi) = \int (fg)\phi\,dx = \int f(g\phi)\,dx = \iota_f(g\phi),$$

so for arbitrary $u \in \mathcal{S}'(\mathbb{R}^n)$ we define $gu \in \mathcal{S}'(\mathbb{R}^n)$ by

$$gu(\phi) = u(g\phi).$$

Note that for $\phi \in \mathcal{S}(\mathbb{R}^n)$, $g\phi \in \mathcal{S}(\mathbb{R}^n)$ since $g \in \mathcal{C}^\infty(\mathbb{R}^n)$ with polynomial bounds, so the definition makes sense.

This actually can be used to provide the first step in the proof of the sequential density of $\mathcal{S}(\mathbb{R}^n)$ in $\mathcal{S}'(\mathbb{R}^n)$. Indeed, if $u \in \mathcal{S}'(\mathbb{R}^n)$, and $\rho \in \mathcal{C}_c^\infty(\mathbb{R}^n)$ is such that $\rho(x) = 1$ for $|x| < 1$, $\rho(x) = 0$ for $|x| > 2$ then $v_j = \rho(x/j)u \to u$ in $\mathcal{S}'(\mathbb{R}^n)$ since this means simply $v_j(\phi) = u(\phi_j) \to u(\phi)$, $\phi_j(x) = \rho(x/j)\phi(x)$, which in turn follows since $\phi_j \to \phi$ in $\mathcal{S}$ by a simple calculation, and since $u : \mathcal{S} \to \mathbb{F}$ is continuous. The gain here is that v_j is compactly supported, namely if $\chi_j \in \mathcal{C}_c^\infty(\mathbb{R}^n)$ is such that $\chi_j(x) = 1$ where $\rho(x/j) \neq 0$, then $\chi_j v_j = v_j$, since this reduces to the statement $(1 - \chi_j(x))\rho(x/j) = 0$ for all x. It then remains to show that such v_j can be approximated in $\mathcal{S}'$ by elements of $\mathcal{C}_c^\infty(\mathbb{R}^n)$. There are different ways of completing the argument; one could even use a Fourier series argument taking advantage of that one could regard a fixed ball as being a compact subset of the interior of a large cube, on which one can use Fourier series. The most straightforward way is, however, convolutions, which we discuss later.

As an example of calculation with distributions, consider the following:

Lemma 0.2. *Suppose that $u \in \mathcal{S}'(\mathbb{R})$ and $xu = 0$. Then there is a constant $c \in \mathbb{F}$ such that $u = c\delta_0$.*

Proof. First note that if $u = c\delta_0$, then $xu = 0$ indeed:

$$xu(\phi) = u(x\phi) = c\delta_0(x\phi) = c(x\phi)(0) = 0$$

for all $\phi \in \mathcal{S}(\mathbb{R})$ since $x(0) = 0$, so $(x\phi)(0) = 0$.

Suppose now that $xu = 0$, i.e. $u(x\phi) = 0$ for all $\phi \in \mathcal{S}(\mathbb{R})$. If ψ is a test function such that $\psi(0) = 0$ then Taylor's theorem allows one to write $\psi = x\phi$ with $\phi \in \mathcal{S}(\mathbb{R})$ (namely $\phi = x^{-1}\psi$ for $x \neq 0$ extends to be $\mathcal{C}^\infty$ at 0), so $u(\psi) = u(x\phi) = 0$.

So now suppose that $\psi \in \mathcal{S}(\mathbb{R})$. Let $\phi_0 \in \mathcal{S}(\mathbb{R})$ be such that $\phi_0(0) \neq 0$. We choose $\alpha \in \mathbb{F}$ such that $\psi - \alpha\phi_0$ vanishes at 0, i.e. let $\alpha = \frac{\psi(0)}{\phi_0(0)}$. Then by the argument of the previous paragraph, $u(\psi - \alpha\phi_0) = 0$. Thus,

$$\begin{aligned} u(\psi) &= u((\psi - \alpha\phi_0) + \alpha\phi_0) = u(\psi - \alpha\phi_0) + \alpha u(\phi_0) \\ &= \alpha u(\phi_0) = \frac{u(\phi_0)}{\phi_0(0)}\psi(0) = c\delta_0(\psi),\ c = \frac{u(\phi_0)}{\phi_0(0)}. \end{aligned}$$

This finishes the proof. □

We defined the Fourier transform on $\mathcal{S}$ as

$$(\mathcal{F}\phi)(\xi) = \hat{\phi}(\xi) = \int_{\mathbb{R}^n} e^{-ix\cdot\xi}\,\phi(x)\,dx,$$

and the inverse Fourier transform as

$$(\mathcal{F}^{-1}\psi)(x) = (2\pi)^{-n}\int_{\mathbb{R}^n} e^{ix\cdot\xi}\,\psi(\xi)\,d\xi.$$

We have shown that the Fourier transform satisfies the relation

$$\int \hat{\phi}(\xi)\psi(\xi)\,d\xi = \int \phi(x)\hat{\psi}(x)\,dx, \qquad \phi, \psi \in \mathcal{S}.$$

In the language of distributional pairing this just says that the tempered distribtions ι_ϕ, resp. $\iota_{\hat{\phi}}$, defined by ϕ, resp. $\hat{\phi}$, satisfy

$$\iota_{\hat{\phi}}(\psi) = \iota_\phi(\hat{\psi}), \qquad \psi \in \mathcal{S}.$$

Motivated by this, we *define* the Fourier transform of an arbitrary tempered distribution $u \in \mathcal{S}'$ by

$$(\mathcal{F}u)(\psi) = u(\hat{\psi}), \qquad \psi \in \mathcal{S}.$$

It is easy to check that $\hat{u} = \mathcal{F}u$ is indeed a tempered distribution, and as observed above, this definition is consistent with the original one if u is a tempered distribution given by a Schwartz function ϕ (or one with enough decay at infinity). It is also easy to see that the Fourier transform, when thus extended to a map $\mathcal{F} : \mathcal{S}' \to \mathcal{S}'$, still has the standard properties, e.g. $\mathcal{F}(D_{x_j} u) = \xi_j \mathcal{F}u$, $D_{x_j} = \frac{1}{i}\partial_{x_j}$. Indeed, by definition, for all $\psi \in \mathcal{S}$,

$$\begin{aligned}(\mathcal{F}(D_{x_j}u))(\psi) &= (D_{x_j}u)(\mathcal{F}\psi) = -u(D_{x_j}\mathcal{F}\psi)\\ &= u(\mathcal{F}(\xi_j\psi)) = (\mathcal{F}u)(\xi_j\psi) = (\xi_j\mathcal{F}u)(\psi),\end{aligned}$$

finishing the proof.

The inverse Fourier transform of a tempered distribution is defined analogously,

$$\mathcal{F}^{-1}u(\psi) = u(\mathcal{F}^{-1}\psi),$$

and it satisfies

$$\mathcal{F}^{-1}\mathcal{F} = \mathrm{Id} = \mathcal{F}\mathcal{F}^{-1}$$

on tempered distributions as well. Again, this is an immediate consequence of the corresponding properties for $\mathcal{S}$, for

$$(\mathcal{F}^{-1}\mathcal{F}u)(\psi) = \mathcal{F}u(\mathcal{F}^{-1}\psi) = u(\mathcal{F}\mathcal{F}^{-1}\psi) = u(\psi).$$

As an example, we find the Fourier transform of the distribution $u = \iota_1$ given by the constant function 1. Namely, for all $\psi \in \mathcal{S}$,

$$\hat{u}(\psi) = u(\hat{\psi}) = \int_{\mathbb{R}^n} \hat{\psi}(x)\,dx = (2\pi)^n\mathcal{F}^{-1}(\hat{\psi})(0) = (2\pi)^n\psi(0) = (2\pi)^n\delta_0(\psi).$$

Here the first equality is from the definition of the Fourier transform of a tempered distribution, the second from the definition of u, the third by realizing that the integral of any function ϕ (in this case $\phi = \hat{\psi}$) is just $(2\pi)^n$ times its inverse Fourier transform evaluated at the origin (directly from the definition of $\mathcal{F}^{-1}$ as an integral), the fourth from $\mathcal{F}^{-1}\mathcal{F} = \mathrm{Id}$ on Schwartz functions, and the last from the definition of the delta distribution. Thus, $\mathcal{F}u = (2\pi)^n\delta_0$, which is often written as $\mathcal{F}1 = (2\pi)^n\delta_0$. Similarly, the Fourier transform of the tempered distribution u given by the function $f(x) = e^{ix\cdot a}$, where $a \in \mathbb{R}^n$ is a fixed constant, is given by $(2\pi)^n\delta_a$ since

$$\hat{u}(\psi) = u(\hat{\psi}) = \int_{\mathbb{R}^n} e^{ix\cdot a}\hat{\psi}(x)\,dx = (2\pi)^n\mathcal{F}^{-1}(\hat{\psi})(a) = (2\pi)^n\psi(a) = (2\pi)^n\delta_a(\psi),$$

while its inverse Fourier transform is given by δ_{-a} since

$$\mathcal{F}^{-1}u(\psi) = u(\mathcal{F}^{-1}\psi) = \int_{\mathbb{R}^n} e^{ix\cdot a}\mathcal{F}^{-1}\psi(x)\,dx = \mathcal{F}(\mathcal{F}^{-1}\psi)(-a) = \psi(-a) = \delta_{-a}(\psi).$$

We can also perform analogous calculations on δ_b, $b \in \mathbb{R}^n$:

$$\mathcal{F}\delta_b(\psi) = \delta_b(\mathcal{F}\psi) = (\mathcal{F}\psi)(b) = \int e^{-ix\cdot b}\psi(x)\,dx,$$

i.e. the Fourier transform of δ_b is the tempered distribution given by the function $f(x) = e^{-ix\cdot b}$. With $b = -a$, the previous calculations confirm what we knew anyway namely that $\mathcal{F}\mathcal{F}^{-1}f = f$ (for this particular f).

If L is a linear partial differential operator, so L is of the form

$$L = \sum_{|\alpha|\leq m} a_\alpha \partial^\alpha,$$

and a_α are in $\mathcal{C}^\infty(\mathbb{R}^n)$ with polynomial bounds for all derivatives, then for all $u \in \mathcal{S}'(\mathbb{R}^n)$, Lu makes sense as an element of $\mathcal{S}'(\mathbb{R}^n)$. In particular, we make the following definition:

Definition 4. Suppose $f \in \mathcal{S}'(\mathbb{R}^n)$, $u \in \mathcal{S}'(\mathbb{R}^n)$, L is linear with $\mathcal{C}^\infty(\mathbb{R}^n)$ coefficients, with polynomial bounds. We say that u is a *weak solution* of $Lu = f$ if $Lu = f$ in the sense of tempered distributions.

Notice that explicitly, with L as above, if both u and f are of the form $(1 + |x|^2)^N L^1(\mathbb{R}^n)$ for some N, this simply means that for all $\phi \in \mathcal{S}(\mathbb{R}^n)$,

$$\sum_{|\alpha|\leq m} \int_{\mathbb{R}^n} u(x)(-1)^{|\alpha|}\partial^\alpha(a_\alpha(x)\phi(x))\,dx = \int_{\mathbb{R}^n} f(x)\phi(x)\,dx,$$

i.e. that

$$\int u(x)L^\dagger\phi(x)\,dx = \int_{\mathbb{R}^n} f(x)\phi(x)\,dx, \tag{2}$$

where $L^\dagger$ is the transpose of L:

$$(L^\dagger\phi)(x) = \sum_{|\alpha|\leq m}(-1)^{|\alpha|}\partial^\alpha(a_\alpha(x)\phi(x)),$$

or simply

$$L^\dagger = \sum_{|\alpha|\leq m}(-1)^{|\alpha|}\partial^\alpha a_\alpha,$$

where a_α are understood as multiplication operators. (So M_{a_α} would be better notation, where $(M_{a_\alpha}\phi)(x) = a_\alpha(x)\phi(x)$, but one usually just abuses the notation and writes a_α for the multiplication operator.)

We can now use tempered distributions to solve the wave equation on $\mathbb{R}^n$. Thus, consider the PDE

$$(\partial_t^2 - c^2\Delta)u = 0,\ u(0,x) = \phi(x),\ u_t(0,x) = \psi(x).$$

Take the partial Fourier transform $\hat{u}$ of u in x to get

$$(\partial_t^2 + c^2|\xi|^2)\hat{u} = 0,\ \hat{u}(0,\xi) = \mathcal{F}\phi(\xi),\ \hat{u}_t(0,\xi) = \mathcal{F}\psi(\xi).$$

For each $\xi \in \mathbb{R}^n$ this is an ODE that is easy to solve, with the result that

$$\hat{u}(t,\xi) = \cos(c|\xi|t)\mathcal{F}\phi(\xi) + \frac{\sin(c|\xi|t)}{c|\xi|}\mathcal{F}\psi(\xi).$$

Thus,

$$u = \mathcal{F}_\xi^{-1}\left(\cos(c|\xi|t)\mathcal{F}\phi(\xi) + \frac{\sin(c|\xi|t)}{c|\xi|}\mathcal{F}\psi(\xi)\right).$$

This can be rewritten in terms of convolutions, namely

$$u(t,x) = \mathcal{F}_\xi^{-1}(\cos(c|\xi|t)) *_x \phi + \mathcal{F}_\xi^{-1}\left(\frac{\sin(c|\xi|t)}{c|\xi|}\right) *_x \psi(\xi),$$

so it remains to evaluate the inverse Fourier transforms of these explicit functions. We only do this in $\mathbb{R}$ (i.e. $n = 1$).

Here we need to be a little careful as we might be taking the convolution of two distributions in principle! However, any tempered distribution can be convolved with elements of $\mathcal{S}(\mathbb{R}^n)$. Indeed, if $f \in C(\mathbb{R}^n)$ polynomially bounded, $\phi \in \mathcal{S}(\mathbb{R}^n)$ then

$$f * \phi(x) = \int f(y)\phi(x-y)\,dy = \iota_f(\phi_x),$$

where we write $\phi_x(y) = \phi(x-y)$. Note that $f*\phi \in \mathcal{C}^\infty(\mathbb{R}^n)$ in fact, as differentiation under the integral sign shows (the derivatives fall on ϕ!). We make the consistent definition for $u \in \mathcal{S}(\mathbb{R}^n)$ that

$$(u * \phi)(x) = u(\phi_x),$$

so $u * \phi \in \mathcal{C}^\infty(\mathbb{R}^n)$ since

$$\partial_{x_j}(u * \phi) = u(\partial_{x_j}\phi_x),$$

in analogy with differentiation under the integral sign, as can be checked by taking difference quotients, and using the fundamental theorem of calculus for $h^{-1}(\phi_{x+h} - \phi_x)$. As an example,

$$\delta_a * \phi(x) = \delta_a(\phi_x) = \phi(x-a).$$

With some work one can even make sense of convolving distributions, as long as one of them has compact support, but we do not pursue this here, as we shall see directly that our formula makes sense for distributions even. We also note that for $f(x) = H(a - |x|)$, $a > 0$, where H is the Heaviside step function (so $H(s) = 1$ for $s \geq 0$, $H(s) = 0$ for $s < 0$),

$$f * \phi(x) = \int H(a - |y|)\phi(x-y)\,dx = \int_{-a}^{a} \phi(x-y)\,dy = \int_{x-a}^{x+a} \phi(s)\,ds,$$

where we wrote $s = x - y$.

Returning to the actual transforms, (using that cos is even, sin is odd)

$$\mathcal{F}_\xi^{-1}(\cos(c\xi t)) = \frac{1}{2}(\mathcal{F}_\xi^{-1}e^{ict\xi} + \mathcal{F}_\xi^{-1}e^{-ict\xi}) = \frac{1}{2}(\delta_{-ct} + \delta_{ct}),$$

while (note that $\xi^{-1}\sin(\xi ct)$ is continuous, indeed $\mathcal{C}^\infty$, at $\xi = 0$!) from the homework

$$\mathcal{F}_x H(ct - |x|) = \frac{2}{\xi}\sin(ct\xi),$$

so

$$\mathcal{F}_\xi^{-1}(c^{-1}\xi^{-1}\sin(c\xi t)) = \frac{1}{2c}H(ct - |x|).$$

In summary,

$$\begin{aligned} u(t,x) &= \frac{1}{2}(\delta_{-ct} + \delta_{ct})) *_x \phi + \frac{1}{2c}H(ct - |x|) *_x \psi \\ &= \frac{1}{2}(\phi(x-ct) + \phi(x+ct)) + \frac{1}{2c}\int_{x-ct}^{x+ct} \psi(y)\,dy, \end{aligned}$$

so we recover d'Alembert's formula.

Finally we comment on the larger picture for distributions. In general, linear PDE theory constructs solutions of PDE by duality arguments. To get a bit of feel for this, consider the following analogue. In finite dimensional vector spaces V, W, if one has a map $P : V \to W$ and $P^* : W^* \to V^*$ is the adjoint, defined by

$$(P^*\ell)(v) = \ell(Pv),\ \ \ell \in W^*,\ \ v \in V,$$

then P is onto if and only if P^* is one-to-one, and P is one-to-one if and only if P^* is onto. (This reduces to a statement about matrices by introducing bases; the adjoint is the transpose, or the conjugate transpose, depending on the definition used.) Thus, if we try to solve $P^*\ell = f$, where $f \in V^*$ is given, we just need to show that P is one-to-one. In infinite dimensions there are serious complications; for one thing instead of just the one-to-one nature, we want an estimate of the form

$$\|v\|_V \leq C\|Pv\|_W,$$

i.e. there is $C > 0$ such that for all $v \in V$ this estimate holds. Notice that this implies the statement that P is injective. The space $\mathcal{S}(\mathbb{R}^n)$ is not normed, which causes some complications; one typically works with a somewhat bigger space instead. In any case, this gives distributional solutions to a PDE in the presence of some estimates for the adjoint operator; in the case of $P^* = L$ above, this amounts to estimates for P which is essentially (up to possibly some complex conjugates in the complex valued setting) $L^\dagger$. (Also, one would typically want to impose extra conditions, such as initial conditions or boundary conditions, so there are further subtleties.)

We also mention that weak solutions give rise to a numerical way of solving PDE. Namely, consider (2). Let's take a finite dimensional space X_N of *trial functions* which we take piecewise continuous, with say a basis $\psi_1, \ldots, \psi_N$, and a finite dimensional space Y_N of *test functions* which we take C^m, with a basis $\phi_1, \ldots, \phi_N$. Let's assume that u can be approximated by a linear combination of the ψ_j, i.e. by $\sum_{j=1}^N c_j \psi_j$. If this combination actually solved the PDE, we would have

$$\int_{\mathbb{R}^n} \big(\sum_{j=1}^N c_j \psi_j(x)\big)(L^\dagger \phi)(x)\, dx = \int_{\mathbb{R}^n} f(x)\phi(x)\, dx,$$

for all $\phi \in \mathcal{C}_c^\infty(\mathbb{R}^n)$. Since we have only finitely many constants c_j to vary, it is unlikely that we can solve the equation so that it is satisfied for all ϕ. However, let us demand that the equation is only satisfied for $\phi \in Y_N$, in which case it suffices to check it for the basis consisting of the ϕ_k, i.e. let us demand

$$\int_{\mathbb{R}^n} \big(\sum_{j=1}^N c_j \psi_j(x)\big)(L^\dagger \phi_k)(x)\, dx = \int_{\mathbb{R}^n} f(x)\phi_k(x)\, dx,\ \ k = 1, \ldots, N.$$

This is N equations ($k = 1, \ldots, N$) for N unknowns (the c_j), so typically (when the corresponding matrix is invertible) we would expect that we can solve these equations, which are just linear equations: they are of the form

$$\sum_{j=1}^N A_{kj} c_j = g_k,$$

where $A_{kj} = \int_{\mathbb{R}^n} \psi_j(x)(L^\dagger \phi_k)(x)\, dx$ and $g_k = \int_{\mathbb{R}^n} f(x)\phi_k(x)\, dx$. Again, there are issues about additional conditions (which would influence the choice of X_N and Y_N), and one may want to write the equations in a different form. For instance for a second order equation have one derivative each on ϕ_k and ψ_j, requiring them to be piecewise C^1. Such a rewriting, with an appropriate choice of ϕ_k and ψ_j is the basis of the very important method of finite elements for numerically solving PDE.