A QUICK DESCRIPTION OF SOME DISTRIBUTIONS (GENERALIZED FUNCTIONS) AND THEIR FOURIER TRANSFORMS

MICHAEL CWIKEL

Version 1.9 (2 February 2006.)

These notes are intended for curious and motivated students who have taken a first course in Fourier transforms, for example as part of the course *Fourier series and integral transforms* at Technion–Israel Institute of Technology.

I have written these notes because several of my colleagues in the Technion's Department of Electrical Engineering asked me whether it might be possible to provide a more solid mathematical foundation for some notions and ideas which play important roles in the courses their students take after learning about Fourier transforms. Well, almost everything is possible, but only if you have enough time (extra hours of lectures) and energy and determination to do it.

As far as I can see, these notes use only material or notions that you have met in the Fourier course or in earlier courses on differential and integral calculus and linear algebra. But they do ask you to think about these notions in some rather new ways.

Our reference for Fourier transforms will be the textbook [PZ] for the above–mentioned course, written by Allan Pinkus and Samy Zafrani. We will use the same notation as in that book. In particular $G(\mathbb{R})$ will denote the set of all functions $f : \mathbb{R} \to \mathbb{C}$ which are piecewise continuous (by which we mean piecewise continuous on each bounded subinterval of $\mathbb{R}$) and which are absolutely integrable on $\mathbb{R}$. For each $f \in G(\mathbb{R})$, the Fourier transform of f is the function $\widehat{f}$ defined by $\widehat{f}(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\omega x} f(x)dx$ for all $\omega \in \mathbb{R}$. We will have to use the following well known facts several times:

$$(1) \qquad \begin{cases} \text{For each } f \in G(\mathbb{R}), \text{ the function } \widehat{f} \text{ is bounded and continuous,} \\ \text{and satisfies } \lim_{\omega\to\pm\infty} \widehat{f}(\omega) = 0. \end{cases}$$

It is obvious that $\widehat{f}$ is bounded. For proofs of the other two properties see e.g. [PZ] Theorem 3.1, p. 96.

The reader can find a list of symbols and some reminders about some relevant basic mathematical definitions for integration etc. at the end of these notes in Sections 6 and 7 respectively.

I warmly thank Professor Adam Shwartz for reading an earlier version of these notes and offering a number of very helpful comments and corrections of misprints. I am also very grateful to Yoni Roll for interesting remarks and corrections of several other misprints. Of course I am responsible for any other errors and misprints which I may have introduced since then.

1. Introduction.

You have probably met the Dirac delta "function" in physics courses. This is the strange object which is usually denoted by $\delta(x)$. It is a "function" which equals 0 at every point $x \neq 0$, but at $x = 0$ its "graph" has an infinitely high "spike". Somehow the "area" under this "spike" which has "height" ∞ and "width" 0 is equal to 1. So we can somehow "integrate" δ. Before I write down its "integral" let us agree here that whenever we are not quite sure that some notation that we want to write is really a well defined mathematical object, we shall warn ourselves of our doubts by writing that notation between quotation marks. So we expect to have "$\int_a^b \delta(x)dx$"$= 1$ for every a and b such that $-\infty \leq a < 0 < b \leq \infty$. If we believe that, then it is perhaps reasonable to suppose that we can define the "product" "$\delta(x)f(x)$" of "$\delta(x)$" with a continuous function $f(x)$ on $[a, b]$ and then the "graph" of this product is 0 everywhere except for a spike of "height" "$f(0) \cdot \infty$" at $x = 0$. If in the "game" we are playing, "$\infty \cdot 0$" $= 1$ then the area of this spike should be $f(0)$ and so "$\int_a^b \delta(x)f(x)dx$"$= f(0)$. But there are all sorts of problems with this "game". For example, how should we define "$\int_a^b \delta(x)f(x)dx$" if f is not continuous at 0? We are used to thinking that if we change the value of a function at one point, for example at 0 then this should not change the value of its integral on an interval including that point. But here this principle is wrong. Also, if we allow ourselves to multiply numbers by ∞ and claim that "$\infty \cdot 0$" $= 1$ then at least one of the associative and commutative laws for multiplication must be wrong. If not we can deduce that **all** numbers are equal. Here is the "proof": For every two numbers p and q,

$$p = p \cdot 1 = \text{“}p \cdot (\infty \cdot 0) = \infty \cdot (0 \cdot p) = \infty \cdot 0 = \infty \cdot (0 \cdot q) = q \cdot (\infty \cdot 0)\text{”} = q \cdot 1 = q.$$

So the delta "function" is *not* a function in the precise sense of the word, and if we assume it is, or use the formula "$\infty \cdot 0 = 1$" carelessly, then we can easily make fools of ourselves and reach incorrect inclusions. Let us nevertheless try to guess some more properties of the delta "function". By what we said before, its "Fourier transform" should be a constant function. We would expect it to be given by

$$\text{“}\widehat{\delta}(\omega)\text{”} = \frac{1}{2\pi} \text{“}\int_{-\infty}^{\infty} \delta(x) e^{-ix\omega} dx\text{”} = \frac{1}{2\pi} \cdot e^{-i \cdot 0 \cdot \omega} = \frac{1}{2\pi}. \tag{2}$$

Then, if the "inverse Fourier theorem" is somehow true in this new setting, it suggests that we have some way to calculate the "integral" "$\int_{-\infty}^{\infty} e^{i\omega x} d\omega$" even though the function $e^{i\omega x}$ is not an integrable function of ω on the interval $(-\infty, \infty)$ for any choice of (constant) x. Perhaps we get

$$\text{“}\int_{-\infty}^{\infty} e^{i\omega x} d\omega\text{”} = 2\pi \text{“}\int_{-\infty}^{\infty} \frac{1}{2\pi} e^{i\omega x} d\omega\text{”} = 2\pi \text{“}\int_{-\infty}^{\infty} \widehat{\delta}(\omega) e^{i\omega x} d\omega\text{”} = 2\pi \text{“}\delta(x)\text{”}.$$

Consider the function $H(t) =$ "$\int_{-\infty}^{t} \delta(x) dx$". This obviously has to be $H(t) = \begin{cases} 0 & , \quad t < 0 \\ 1 & , \quad t > 0 \end{cases}$. But what is $H(0)$? Except for our embarrassment about deciding the value at 0, this function H is the "Heaviside function". (It is denoted by u or u_0 in the chapter of [PZ] dealing with Laplace transforms.) If $H(t) =$ "$\int_{-\infty}^{t} \delta(x) dx$" then it seems reasonable to guess that H somehow has a "derivative", even at 0, despite its discontinuity at 0, and that "derivative" should be "$H' = \delta$". Can we also "differentiate" "δ" itself? What is its derivative?

Starting in the late 1800's a number of physicists and mathematicians played with these sorts of ideas and wondered about questions like the last one. There were many intuitive calculations, and good reasons, sometimes from physics, to believe them, and other good reasons (like our comments above) to doubt them.

As time passed, ways were found to treat the delta "function" and other related objects in a precise way. The most successful and now most widely used way of doing this was developed by the great French mathematician Laurent Schwartz beginning with his initial work in 1944[1]. The development of this topic is a particularly good example of the way science can and should develop. On the one hand we should not be frightened to try to work with intuitive ideas, even if at first they seem doubtful or even partly crazy. On the other hand there should also be a parallel process of carefully examining and verifying these ideas, and seeing if they can be expressed in new more precise and more rigorous ways.

In these few pages we can only give a quick glimpse of some part of Laurent Schwartz' work. He introduced a family of new mathematical objects which he called *distributions*. They are sometimes also called *generalized functions*. In particular, for working with the Fourier transform, he introduced a special collection of distributions which he called *tempered distributions*. These form a vector space[2] which is usually denoted by $\mathcal{S}'$. All the functions in $G(\mathbb{R})$ are in $\mathcal{S}'$ and many other functions, constants, all polynomials and many other functions which are not integrable on $\mathbb{R}$ are in $\mathcal{S}'$. But many of the elements in $\mathcal{S}'$ are not functions. In particular, there is an *exact way* of defining the delta "function" as an element of $\mathcal{S}'$.

The space $\mathcal{S}'$ has many remarkable properties. For now we will mention only two of them:

1. It turns out to be possible to define the Fourier transform of *every* element of $\mathcal{S}'$. If the element happens to be a function in $G(\mathbb{R})$ then the new definition and the old definition of Fourier transform give the same thing. Otherwise the new definition may sometimes give a function, (for example $\widehat{\delta}$ is a constant, just as we guessed above) and sometimes it gives something which is not a function. But the Fourier transform of any element of $\mathcal{S}'$ is always an element of $\mathcal{S}'$.

---

[1]Six years later this discovery earned him the Fields Medal (generally considered to be the equivalent of the Nobel Prize for mathematicians).

[2]All vector spaces in these notes are in fact vector spaces over the complex field. For the relevant definition and further comments see Section 7.

2. It is also possible to define the derivative of every element in $\mathcal{S}'$. If the element happens to be a differentiable function in the usual sense of the word and it and its derivative also satisfy some mild "growth" conditions, then the new and old definition of derivative coincide. If not, then the derivative may fail to be a function, but it will always be an element of $\mathcal{S}'$. In particular we can show that our guess above that "$H' = \delta$" is correct, and we can give this equation an exact meaning. We can define the nth derivative of δ (or of any other tempered distribution) for every $n \in \mathbb{N}$.

Before we can define $\mathcal{S}'$ and describe how we define derivatives and Fourier transforms of its elements, we need quite a number of preliminary observations and results.

Suppose $f : \mathbb{R} \to \mathbb{R}$ is a function which arises in the "real world". It could, for example be a function of time representing, for example, an audio signal, say a bird singing. Hopefully we have a good microphone and other equipment which will enable us to see a good approximation to the graph of f on the screen of an oscilloscope or of a computer. But can we exactly measure or know the value of $f(t)$ at each instant t? Apparently we cannot. Instead our equipment in fact measures (approximately) the averages of f over very short periods of time, i.e. the quantities $\frac{1}{b-a}\int_a^b f(x)dx$ for very short intervals $[a, b]$. If f is a continuous function then

$$f(t) = \lim_{h\to 0} \frac{1}{2h}\int_{t-h}^{t+h} f(x)dx. \tag{3}$$

So we can get an approximation to the value of $f(t)$ for some fixed t by using the averages of f on smaller and smaller intervals containing t.

Of course our equipment will not allow us to take the intervals smaller than some strictly positive number (which depends on the amount of money we paid for our equipment and what was its year of manufacture). But let us now leave the birds and the "real world" and go back to thinking more mathematically. The previous remarks suggest that instead of studying a function $f : \mathbb{R} \to \mathbb{C}$ directly, we can try to study it indirectly via the numbers $\int_a^b f(x)dx$ for all values of a and b. But do these numbers contain *all* the information about the function? Yes, they do, at least when f is continuous. This follows immediately from (3). A closely related observation is contained in the next theorem.

Theorem 1.1. *Suppose that $f : \mathbb{R} \to \mathbb{C}$ and $g : \mathbb{R} \to \mathbb{C}$ are two continuous functions which satisfy*

$$\int_a^b f(x)dx = \int_a^b g(x)dx \text{ for all numbers } a \text{ and } b \text{ such that } a < b. \tag{4}$$

Then $f(x) = g(x)$ for all $x \in \mathbb{R}$.

Proof. This follows immediately from (3). ■

If f and g are only piecewise continuous[3] then we get a conclusion which slightly weaker than in Theorem 1.1. The condition (4) implies that $f(x) = g(x)$ at every point $x \in \mathbb{R}$ where f and g are both continuous. So f and g are "almost" the same function. They can differ at most on a "small" set of points E which only has finitely many elements in any bounded interval. So, for example, if f and g both happen to be in $G(\mathbb{R})$ then they both have the same Fourier transform. If we want to be able to conclude that $f(x) = g(x)$ at *all* points, including the points where they have "jumps" (points where their left and right one–sided limits are different), then we can decide, for example, to only work with piecewise continuous functions f which satisfy the extra condition

$$f(x) = \frac{1}{2}\left(f(x+) + f(x-)\right) \tag{5}$$

at each point of discontinuity x. Of course (5) is also true at all other points. It is not hard to see that the formula (3) is true for *all* $t \in \mathbb{R}$ for functions f satisfying (5) for each $x \in \mathbb{R}$. It will be convenient to have a name for the collection of all functions $f : \mathbb{R} \to \mathbb{C}$ which are piecewise continuous on every bounded subinterval of $\mathbb{R}$ and also satisfy (5) for all $x \in \mathbb{R}$. Let us call it PC_{bj}. (The letters come from "Piecewise Continuous with Balanced Jumps".) PC_{bj} is of course a vector space.

[3] The definition of piecewise continuity is recalled in Section 7. Here we are in fact only assuming that f is piecewise continuous on every bounded subinterval of $\mathbb{R}$.

Let us rewrite this variant of the previous theorem in a slightly different way. Suppose that A is the space PC_{bj} and B is the set of all characteristic functions [4] $\chi_{[a,b]}$ of intervals $[a,b]$ for each constant a and b with $-\infty < a < b < \infty$.

$$\left\{ \begin{array}{l} \text{If } f \text{ and } g \text{ are in } A \text{ and} \\ \int_{-\infty}^{\infty} f(x)\phi(x)dx = \int_{-\infty}^{\infty} g(x)\phi(x)dx \textit{ for all functions } \phi \textit{ in } B, \\ \text{then } f(x) = g(x) \text{ for all } x \in \mathbb{R}. \end{array} \right. \tag{6}$$

We now want to consider some other rather different examples of pairs of sets of functions A and B which have the same property (6). In general, if A and B are sets of functions on $\mathbb{R}$, we will say that B *is a separating class for* A if, for every $f \in A$ and every $\phi \in B$ the function $f(x)\phi(x)$ is in $G(\mathbb{R})$ and condition (6) holds. We can abbreviate this terminology and say simply that B *separates* A. The good reason behind this terminology is that, if f and g are two different functions in A, then there is at least one element ϕ in B which "separates" them, i.e. it satisfies $\int_{-\infty}^{\infty} f(x)\phi(x)dx \neq \int_{-\infty}^{\infty} g(x)\phi(x)dx$. (Many mathematicians use another alternative terminology here and say that B *is a total set for* A.)

The functions in a separating class B are sometimes called "test functions" (*funktsiyot bdikah*), because in order to know if two functions f and g in A are equal it is enough to test the behaviour of their integrals with all the functions ϕ in B.

Example 1.2. *Let $A = G(\mathbb{R}) \cap PC_{bj}$ and let B be the set of functions $\phi : \mathbb{R} \to \mathbb{C}$ of the form $\phi(x) = e^{icx}$ for some real constant c. Then B is a separating class for A. To prove this, suppose that f and g in A satisfy $\int_{-\infty}^{\infty} f(x)\phi(x)dx = \int_{-\infty}^{\infty} g(x)\phi(x)dx$ for all ϕ in B, i.e. $\int_{-\infty}^{\infty} f(x)e^{icx}dx = \int_{-\infty}^{\infty} g(x)e^{icx}dx$ for all $c \in \mathbb{R}$. So $\widehat{f}(-c) = \widehat{g}(-c)$ for all $c \in \mathbb{R}$, i.e. $\widehat{f}$ and $\widehat{g}$ are the same function. If f and g have left and right derivatives at every point we can apply the inverse Fourier theorem, i.e. the fact (cf. [PZ] Theorem 3.3, p. 109) that*

$$\begin{array}{l} \lim_{R\to\infty} \int_{-R}^{R} e^{ix\omega}\widehat{u}(x)dx = \frac{1}{2}\left(u(x+) + u(x-)\right) \\ \textit{if } u \in G(\mathbb{R}) \textit{ and } u \textit{ has right and left derivatives at } x. \end{array} \tag{7}$$

to deduce that $f(x) = g(x)$ for all $x \in \mathbb{R}$. (Note that here we have used the condition (5).) If $\lim_{R\to\infty} \int_{-R}^{R} |f(x) - g(x)|^2 dx$ is finite, then we can apply Plancherel's theorem to show that $\frac{1}{2\pi}\int_{-\infty}^{\infty} |f(x) - g(x)|^2 dx = \int_{-\infty}^{\infty} \left|\widehat{f}(\omega) - \widehat{g}(\omega)\right|^2 d\omega = 0$ and this again will give us that $f(x) = g(x)$ for all $x \in \mathbb{R}$. If neither of these conditions hold then we need a more complicated proof, which I will not write here, but which leads to the same conclusion.

Example 1.3. *Let A be the set of functions f in PC_{bj} with the additional property that, for some positive constant $C > 0$ and some positive integer N,*

$$|f(x)| \le C\left(1 + |x|\right)^N \textit{ for all } x \in \mathbb{R}. \tag{8}$$

(We stress that the numbers C and N will be different for different functions f in A.) Let B be the set of all functions $\phi : \mathbb{R} \to \mathbb{C}$ of the form $\phi(x) = e^{-i\omega x} \cdot e^{-x^2}$ for some constant $\omega \in \mathbb{R}$. Again the constant ω takes different values for different functions ϕ in B. Let us try to compare this example with Theorem 1.1. There we were trying to get all information about a function, or audio signal, by looking at all its averages on all "sharp windows". Here we are trying to get all information about a function, or audio signal, by looking at all the averages of its "frequencies" (i.e. Fourier transform) on all translations of a certain "smooth window".

We will now show that also in this case B is a separating class for A. First, since $\lim_{|x|\to\infty} e^{-x^2}(1+|x|)^{n+2} = 0$ for every constant $n \in \mathbb{N}$, it follows that $K_n := \sup_{x\in\mathbb{R}} e^{-x^2}(1+|x|)^{n+2}$ is finite for each $n \in \mathbb{N}$. Consequently, if f is any function in A, then it follows from (8) that, for any $\phi(x) = e^{-i\omega x}e^{-x^2}$,

$$\begin{aligned} |f(x)\phi(x)| &= \left|f(x)e^{-i\omega x}e^{-x^2}\right| \le C \cdot \frac{e^{-x^2}(1 + |x|)^{N+2}}{(1 + |x|)^2} \\ &\le C \cdot K_N \cdot \frac{1}{(1 + |x|)^2} \textit{ for all } x \in \mathbb{R}. \end{aligned} \tag{9}$$

[4] The function $\chi_{[a,b]} : \mathbb{R} \to \mathbb{R}$ is defined by $\chi_{[a,b]} = \left\{ \begin{array}{ll} 1 & \text{if } x \in [a,b] \\ 0 & \text{if } x \notin [a,b] \end{array} \right.$. See also Section 6.

This shows that $f\phi$ is absolutely integrable and of course it is piecewise continuous and satisfies (5) for all x. Now suppose that f and g are any two functions in A which satisfy $\int_{-\infty}^{\infty} f(x)\phi(x)dx = \int_{-\infty}^{\infty} g(x)\phi(x)dx$ for all ϕ in B. Let us define two new functions $F(x) = f(x)e^{-x^2}$ and $G(x) = g(x)e^{-x^2}$. Then the previous condition becomes $\int_{-\infty}^{\infty} F(x)e^{-i\omega x}dx = \int_{-\infty}^{\infty} G(x)e^{-i\omega x}dx$ for all $\omega \in \mathbb{R}$. By estimates similar to (9), the functions F and G are both in $G(\mathbb{R})$ and so the same arguments as in Example 1.2 show that $F = G$ and, consequently, $f = g$. So, indeed, B separates A.

The set A is "very big" but it does not contain all functions in $G(\mathbb{R}) \cap PC_{bj}$.[5] We would like to replace A by a still bigger class of functions which contains all the functions of $G(\mathbb{R}) \cap PC_{bj}$. This bigger class will be denoted by $SG(\mathbb{R})$. The letters SG stand for "slow growth" and the functions in this class are sometimes called "functions of slow growth". In various books the class of functions of slow growth is defined differently, and the notation is also different. We have chosen a definition to suit our particular modest purposes here[6].

Definition 1.4. *We define $SG(\mathbb{R})$ to be the set of all functions in PC_{bj} which satisfy, for some constants $C > 0$ and $N \in \mathbb{N}$,*

$$\int_{-R}^{R} |f(x)|\, dx \le C(1+R)^N \text{ for all } R > 0. \tag{10}$$

Here again the constants C and N will be different for different functions in $SG(\mathbb{R})$.

Obviously $SG(\mathbb{R})$ contains $G(\mathbb{R}) \cap PC_{bj}$.

Example 1.5. *As in Example 1.3, we let B be the set of all functions of the form $\phi(x) = e^{-i\omega x} \cdot e^{-x^2}$ for some constant $\omega \in \mathbb{R}$.Then B separates $SG(\mathbb{R})$. The main step for proving this is to show that, for each f in $SG(\mathbb{R})$, the function $f(x)e^{-x^2}$ is absolutely integrable. This is not difficult to do, but I ask you to believe it for the moment. It will follow immediately from the properties 2.2 and 2.5 which we will soon prove. After we know that $f(x)e^{-x^2}$ is integrable, the rest of the proof that B separates $S(\mathbb{R})$ uses what we know about Fourier transforms of functions in $G(\mathbb{R})$ in exactly the same way as was done in Example 1.3.*

2. The Schwartz class $\mathcal{S}$ and some of its properties.

Our next step will be to replace the set of functions B which appeared in the last two examples by a larger set of functions. This is a family denoted by $\mathcal{S}$ and sometimes called *the Schwartz class*, or sometimes *the class of C^∞ rapidly decreasing functions.*

Definition 2.1. *The class $\mathcal{S}$ consists of all functions $\phi : \mathbb{R} \to \mathbb{C}$ such that the derivative of order n, $\phi^{(n)}(x)$ exists for every $n \in \mathbb{N}$ and every $x \in \mathbb{R}$ and which satisfy the condition*

$$\lim_{x \to \pm\infty} x^m \phi^{(n)}(x) = 0 \text{ for every pair of fixed integers } m \ge 0 \text{ and } n \ge 0. \tag{11}$$

It will turn out that $\mathcal{S}$ plays a central role in defining Laurent Schwartz' tempered distributions. Let us now establish a number of useful properties of $\mathcal{S}$. We would perhaps not expect these properties of the very nice, very smooth, very quickly decaying functions in $\mathcal{S}$ to have any connection with treating quite nasty "functions" which are not even really functions. But we will soon see that there is, *davka*, a very strong connection:

- **2.2.** *Each function ϕ of the form $\phi(x) = e^{-i\omega x} \cdot e^{-x^2}$ for some constant $\omega \in \mathbb{R}$ is in $\mathcal{S}$.*

This is easy to see because for each such function ϕ, and for each $n \in \mathbb{N}$, $\phi^{(n)}(x) = p_n(x) \cdot e^{-i\omega x} \cdot e^{-x^2}$ for some polynomial p_n. This can easily be proved by induction on n. Since $\lim_{x\to\pm\infty} x^k e^{-x^2} = 0$ for all $k \ge 0$, we immediately deduce that (11) holds for all non negative integers m and n.

---

[5] Consider for example the following function f. We set $f(x) = 0$ except on each of the intervals $I_n = \left(n - \frac{1}{2^n n^2}, n + \frac{1}{2^n n^2}\right)$ for all $n = 2, 3, \ldots$. On I_n set $f(n) = 2^n$ and extend f linearly on $\left[n - \frac{1}{2^n n^2}, n\right]$ and on $\left[n, n + \frac{1}{2^n n^2}\right]$. Since $f\left(n \pm \frac{1}{2^n n^2}\right) = 0$ this means that $f \ge 0$ the region under graph of f on I_n is a triangle of area $\frac{1}{2} \cdot \frac{2}{2^n n^2} 2^n = \frac{1}{n^2}$. It is easy to see that this function f is in $G(\mathbb{R})$ and it is even continuous, but it is not in A.

[6] The reasons for choosing this definition are that, for simplicity, we do not want to use functions which are not piecewise continuous and we do not want to use the Lebesgue integral. The Lebesgue integral is a very interesting and powerful generalization of the Riemann integral, but its definition is too complicated to be discussed in these notes or in the course "Fourier series and integral transforms".

• **2.3.** *$\mathcal{S}$ separates $SG(\mathbb{R})$.*

This follows from property 2.2 and Example 1.5.

• **2.4.** *Any infinitely differentiable function $\phi : \mathbb{R} \to \mathbb{C}$ is in $\mathcal{S}$ if and only if it satisfies*

$$\text{(12)} \qquad \begin{array}{l} \textit{For each pair of fixed integers } m \geq 0 \textit{ and } n \geq 0 \textit{ there exists} \\ \textit{a constant } C_{m,n}(\phi) \textit{ such that } \left|x^m \phi^{(n)}(x)\right| \leq C_{m,n}(\phi) \textit{ for all } x \in \mathbb{R}. \end{array}$$

Since a continuous function on $\mathbb{R}$ which has finite limits at $-\infty$ and ∞ must be bounded, it is easy to see that (11) implies (12). Conversely, (12) implies that $\left|x^m \phi^{(n)}(x)\right| = \frac{1}{|x|}\left|x^{m+1}\phi^{(n)}(x)\right| \leq \frac{1}{|x|}C_{m+1,n}(\phi)$ for all $x \neq 0$ and all non negative integers m and n. This in turn implies (11).

The constants $C_{m,n}(\phi)$ in (12) can be chosen to be the numbers $\sup_{x\in\mathbb{R}}\left|x^m\phi^{(n)}(x)\right|$. It is not hard to see that this supremum is also a maximum. So from here onwards, for each $\phi \in \mathcal{S}$, we shall always use the notation

$$C_{m,n}(\phi) = \sup\left\{\left|x^m\phi^{(n)}(x)\right| : x \in \mathbb{R}\right\} = \max\left\{\left|x^m\phi^{(n)}(x)\right| : x \in \mathbb{R}\right\}. \tag{13}$$

• **2.5.** *For each $f \in SG(\mathbb{R})$ and each $\phi \in \mathcal{S}$, the function $f\phi$ is absolutely integrable on $\mathbb{R}$.*

You might think that in the proof of property 2.5 which we will give now, we are worrying too much about the constants in the inequalities. But this is because we will need these inequalities for another purpose later:

Choose an arbitrary function $f \in SG(\mathbb{R})$ and let N and C be the constants appearing in (10). Then $|\phi(x)| \leq C_{0,0}(\phi)$ and $|x^{N+2}\phi(x)| \leq C_{N+2,0}(\phi)$ and so

$$\begin{aligned} \int_{-\infty}^{\infty} |f(x)\phi(x)|\,dx &= \int_{-1}^{1} |f(x)\phi(x)|\,dx + \left(\int_{-\infty}^{-1} |f(x)\phi(x)|\,dx + \int_{1}^{\infty} |f(x)\phi(x)|\,dx\right) \\ &\leq C_{0,0}(\phi)\int_{-1}^{1}|f(x)|\,dx + C_{N+2,0}(\phi)\left(\int_{-\infty}^{-1}\frac{|f(x)|}{|x|^{N+2}}dx + \int_{1}^{\infty}\frac{|f(x)|}{|x|^{N+2}}dx\right). \end{aligned}$$

By (10) we have $\int_{-1}^{1}|f(x)|\,dx \leq C \cdot 2^N$. We also have

$$\begin{aligned} \int_{-\infty}^{-1}\frac{|f(x)|}{|x|^{N+2}}dx + \int_{1}^{\infty}\frac{|f(x)|}{|x|^{N+2}}dx &= \sum_{n=1}^{\infty}\left(\int_{-n-1}^{-n}\frac{|f(x)|}{|x|^{N+2}}dx + \int_{n}^{n+1}\frac{|f(x)|}{|x|^{N+2}}dx\right) \\ &\leq \sum_{n=1}^{\infty}\frac{1}{n^{N+2}}\left(\int_{-n-1}^{-n}|f(x)|dx + \int_{n}^{n+1}|f(x)|dx\right) \\ &\leq \sum_{n=1}^{\infty}\frac{1}{n^{N+2}}\int_{-n-1}^{n+1}|f(x)|\,dx. \end{aligned}$$

Using (10) this last expression is dominated by

$$\sum_{n=1}^{\infty}\frac{1}{n^{N+2}} \cdot C(1+n+1)^N = C \cdot \sum_{n=1}^{\infty}\frac{(n+2)^N}{n^{N+2}}.$$

The general term in this series satisfies

$$\frac{(n+2)^N}{n^{N+2}} = \frac{(n+2)^N}{n^N}\cdot\frac{1}{n^2} = \left(1+\frac{2}{n}\right)^N\frac{1}{n^2} \leq 3^N\frac{1}{n^2}. \tag{14}$$

Combining all the preceding estimates, we obtain that

$$\int_{-\infty}^{\infty}|f(x)\phi(x)|\,dx \leq C \cdot \left(2^N C_{0,0}(\phi) + C_{N+2,0}(\phi)\cdot 3^N \cdot \sum_{n=1}^{\infty}\frac{1}{n^2}\right). \tag{15}$$

Please remember that here N is a CONSTANT, and n is the variable of summation. Since the series $\sum_{n=1}^{\infty}\frac{1}{n^2}$ converges, the estimate (15) shows that $f\phi$ is indeed absolutely integrable on $(-\infty, \infty)$. (Note that by proving this we have now also established the claim made in Example 1.5.)

• **2.6.** *$\mathcal{S}$ is a vector space.*

It is obvious that whenever ϕ and ψ are functions in $\mathcal{S}$, then $\alpha\phi+\beta\psi$ is a function in $\mathcal{S}$ for all complex constants α and β. It is very easy to verify the remaining conditions needed to show that $\mathcal{S}$ is a vector space. (Cf. the discussion about vector spaces in Section 7.)

• **2.7.** *For each $\phi\in\mathcal{S}$, each integer $n\geq 0$ and each polynomial p, the function $p(x)\phi^{(n)}(x)$ is bounded and absolutely integrable and is also in $\mathcal{S}$.*

It follows immediately from the definitions and condition (12), that $p(x)\phi^{(n)}(x)$ is in $\mathcal{S}$ and is bounded. For exactly the same reasons $(1+x^2)p(x)\phi^{(n)}(x)$ is, among other things, bounded in absolute value by some constant C. Consequently $\left|p(x)\phi^{(n)}(x)\right|\leq\frac{C}{1+x^2}$ showing that $p(x)\phi^{(n)}(x)$ is absolutely integrable.

• **2.8.** *Whenever ϕ is in $\mathcal{S}$, its Fourier transform $\widehat{\phi}$ is also in $\mathcal{S}$.*

Since $\phi\in G(\mathbb{R})$ it follows (cf. (1)) that $\widehat{\phi}$ is bounded and continuous. By property 2.7 the function $\psi(x)=-ix\phi(x)$ is also in $G(\mathbb{R})$ and so its Fourier transform $\widehat{\psi}$ is bounded and continuous. So we can then apply another standard result, involving differentiation through the integral sign (with the help of the Lebesgue dominated convergence theorem), to obtain that $\widehat{\phi}$ is differentiable and its derivative is $\widehat{\psi}$, i.e. $\frac{d}{d\omega}\widehat{\phi}(\omega)=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-i\omega x}(-ix)\phi(x)dx$. Since $\psi\in\mathcal{S}$ we can repeat the same argument to show that $\widehat{\psi}'=\widehat{\phi}''$ is continuous and bounded and is the Fourier transform of yet another function in $\mathcal{S}$. In fact we can repeat this argument as often as we wish, and obtain that:

$$\text{For each } n\in\mathbb{N}, \frac{d^n}{d\omega^n}\widehat{\phi}(\omega) \text{ exists and is the Fourier transform of some function in } \mathcal{S}. \tag{16}$$

More precisely

$$\frac{d^n}{d\omega^n}\widehat{\phi}(\omega)=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-i\omega x}(-ix)^n\phi(x)dx. \tag{17}$$

Now we will recall and use another standard result: Let ψ be a differentiable function in $G(\mathbb{R})$ such that ψ' is also in $G(\mathbb{R})$. Then

$$\widehat{\psi'}(\omega)=i\omega\widehat{\psi}(\omega). \tag{18}$$

The first step towards obtaining (18) is to observe that, for any interval $[a,b]$, we have

$$\left(e^{-i\omega x}\psi(x)\right)\Big|_{x=a}^{b}=\int_a^b\frac{d}{dx}\left[e^{-i\omega x}\psi(x)\right]dx=\int_a^b -i\omega e^{-i\omega x}\psi(x)dx+\int_a^b e^{-i\omega x}\psi'(x)dx. \tag{19}$$

The rest of the proof of (18) is almost immediate when we also know that $\lim\limits_{x\to\pm\infty}\psi(x)=0$. In fact here we will be assuming much more, i.e. that $\psi\in\mathcal{S}$. Since $\lim\limits_{x\to\pm\infty}\psi(x)=0$ we can let a tend to $-\infty$ and b tend to $+\infty$ in (19) and divide by 2π to obtain (18). The importance of this formula for us now is that it shows that whenever $\psi\in\mathcal{S}$, then $\omega\widehat{\psi}(x)$ is the Fourier transform of another function in $\mathcal{S}$. Repeating this argument m times for any integer $m\in\mathbb{N}$, we can see that $\omega^m\widehat{\psi}(x)$ is also the Fourier transform of a function in $\mathcal{S}$. In particular, since $\mathcal{S}\subset G(\mathbb{R})$ this means that

$$\lim_{\omega\to\pm\infty}\omega^m\widehat{\psi}(\omega)=0 \text{ for each integer } m\geq 0. \tag{20}$$

For later use let us also note more explicitly that in this way we obtain (cf. (18)) that

$$\widehat{\psi^{(m)}}(\omega)=(i\omega)^m\widehat{\psi}(\omega) \text{ for each } \psi\in\mathcal{S} \text{ and } m\in\mathbb{N}. \tag{21}$$

Now, given any $\phi\in\mathcal{S}$ and any integer $n\geq 0$, let us choose ψ to be the function in $\mathcal{S}$ which exists in view of (16) with the property that $\widehat{\psi}(\omega)=\frac{d^n}{d\omega^n}\widehat{\phi}(\omega)$. Then, by (20) we have that $\lim_{\omega\to\pm\infty}\omega^m\frac{d^n}{d\omega^n}\widehat{\phi}(\omega)=0$. In other words, $\widehat{\phi}\in\mathcal{S}$, and we have proved property 2.8.

• **2.9.** *For each interval $[a,b]$ there exists a function $\phi\in\mathcal{S}$ such that $\phi(x)>0$ for all $x\in(a,b)$ and $\phi(x)=0$ for all $x\notin(a,b)$.*

The construction of such functions is not completely obvious. For details, see Subsection 5.1 below.

3. The space $\mathcal{S}'$ of tempered distributions.

We are now ready, or almost ready, I hope, to make the big jump from the idea of a function to the idea of a generalized function. Examples 1.2,1.3 and 1.5 and property 2.3 of $\mathcal{S}$ tell us that to determine a function f in the space $G(\mathbb{R}) \cap PC_{bj}$ or even in the much bigger space $SG(\mathbb{R})$, it is enough to know the values of the integrals $\int_{-\infty}^{\infty} f(x)\phi(x)dx$ for all or some of the "test functions" $\phi \in \mathcal{S}$. So instead of thinking about the values $f(x)$ of f at each point $x \in \mathbb{R}$ we can equivalently think about the values $L_f(\phi)$ of a new "function", which we will call L_f, at each "point" ϕ in $\mathcal{S}$. These values are given by the formula

$$L_f(\phi) := \int_{-\infty}^{\infty} f(x)\phi(x)dx \text{ for all } \phi \in \mathcal{S}. \tag{22}$$

The generalized functions, or tempered distributions of Laurent Schwartz will also be "functions" L defined on $\mathcal{S}$ instead of on $\mathbb{R}$. I.e., for each $\phi \in \mathcal{S}$ we have a complex number $L(\phi)$. In other words L is a map from $\mathcal{S}$ to $\mathbb{C}$. In general $L(\phi)$ will *not* be given by a formula like (22). But L will be required to have some special properties:

Definition 3.1. *Let L be a map from the space $\mathcal{S}$ to the complex numbers which satisfies the following two conditions:*

(Linearity): For each ϕ and ψ in $\mathcal{S}$ and each α and β in $\mathbb{C}$

$$L(\alpha\phi + \beta\psi) = \alpha L(\phi) + \beta L(\psi).$$

(A special kind of continuity): If $\{\phi_k\}_{k\in\mathbb{N}}$ is a sequence of functions in $\mathcal{S}$ such that, for each pair of constant integers $m \geq 0$ and $n \geq 0$, the numbers $C_{m,n}(\phi_k)$ defined by (13) satisfy $\lim_{k\to\infty} C_{m,n}(\phi_k) = 0$, then

$$\lim_{k\to\infty} L(\phi_k) = 0. \tag{23}$$

Then L is said to be a ***tempered distribution****.*

The set of all maps L with these properties is denoted by $\mathcal{S}'$.

Before we show that the delta "function" and other generalized "functions" can be obtained as elements of $\mathcal{S}'$ let us see that $\mathcal{S}'$ "contains" all the functions of $SG(\mathbb{R})$. More precisely, we claim that, for each function $f \in SG(\mathbb{R})$, the linear map L_f defined by the formula (22) is in $\mathcal{S}'$. The map L_f obviously satisfies the linearity condition of Defintion 3.1. To show that the continuity condition also holds we use the estimate (15) which immediately gives that, for every $\phi \in \mathcal{S}$,

$$|L_f(\phi)| \leq \int_{-\infty}^{\infty} |f(x)\phi(x)|\, dx \leq C \cdot \left(2^N C_{0,0}(\phi) + C_{N+2,0}(\phi) \cdot 3^N \cdot K\right),$$

where K is the constant $K = \sum_{n=1}^{\infty} \frac{1}{n^2}$, (it equals $\pi^2/6$ but the exact value is not important here) and C and N are also constants which only depend on our choice of f. This estimate shows that if $C_{0,0}(\phi_k)$ and $C_{N+2,0}(\phi_k)$ both tend to 0 as k tends to ∞, then (23) holds.

The fact that $\mathcal{S}$ separates $SG(\mathbb{R})$ shows that for each $f \in SG(\mathbb{R})$ there is only one map L in $\mathcal{S}'$ such that $L = L_f$. In fact we can also calculate the value of $f(x)$ for each $x \in \mathbb{R}$ if we know the value of $L_f(\phi)$ for all $\phi \in \mathcal{S}$. We will give an exact formula for doing this in an appendix (Subsection 5.2). So we have a well defined one to one correspondence between functions $f \in SG(\mathbb{R})$ and the maps L_f in $\mathcal{S}'$. If we identify f and L_f we can think of $SG(\mathbb{R})$ as a subset of $\mathcal{S}'$.

Now let us consider the map $\delta_0 : \mathcal{S} \to \mathbb{C}$ defined by $\delta_0(\phi) = \phi(0)$ for each $\phi \in \mathcal{S}$. This map is obviously linear: $\delta_0(\alpha\phi + \beta\psi) = \alpha\phi(0) + \beta\psi(0) = \alpha\delta_0(\phi) + \beta\delta_0(\psi)$. It also satisfies the second condition of Definition 3.1 since, obviously, $|\delta_0(\phi)| \leq C_{0,0}(\phi)$. So δ_0 is a tempered distribution. It is not very difficult to show that there is **no** function f in $SG(\mathbb{R})$ such that $\delta_0(\phi) = \int_{-\infty}^{\infty} f(x)\phi(x)dx$. Again we will defer the exact proof of this to Section 5 (See Subsection 5.3). The map δ_0 is a well defined map which plays the role of the delta "function" in an exact way. We can also define translates of δ_0. For each constant $a \in \mathbb{R}$, let δ_a be the map from $\mathcal{S}$ to $\mathbb{C}$ defined by $\delta_a(\phi) = \phi(a)$. This too is a tempered distribution and it is the well defined version of what is sometimes denoted by "$\delta(x-a)$".

I said above that we can find the derivative of the delta "function". That seems crazy at first, since the delta "function" is not even continuous, and it is not even a function. But if the derivative also does not have to be a function, then maybe there is hope of finding it.

The first step towards defining derivatives of tempered distributions is to recall the formula for integration by parts

$$\int_a^b f'(x)\phi(x)dx = \left(f(x)\phi(x)\right)\Big|_{x=a}^b - \int_a^b f(x)\phi'(x)dx \tag{24}$$

which holds whenever f and ϕ are both continuous functions on the interval $[a,b]$ whose derivatives f' and ϕ' exist at every point of (a,b) and are both Riemann integrable functions on (a,b). In particular (24) holds for every real a and b if f is any function in $SG(\mathbb{R})$ which is differentiable at every point of $\mathbb{R}$ and whose derivative f' is also in $SG(\mathbb{R})$ and if ϕ is any function in $\mathcal{S}$.

We want to choose arbitrary f and ϕ with these properties, and take the limit in (24) as a tends to $-\infty$ and b tends to $+\infty$. The problem is that we cannot be sure in advance that $\lim_{b\to\infty} f(b)\phi(b)$ and $\lim_{a\to-\infty} f(a)\phi(a)$ exist. If we impose some extra condition, such as requiring f to also be in the class A discussed in Example 1.3, then both these limits do exist and are 0 and so we obtain that

$$\int_{-\infty}^{\infty} f'(x)\phi(x)dx = -\int_{-\infty}^{\infty} f(x)\phi'(x)dx. \tag{25}$$

In fact with a little bit more care (see Subsection 5.4) we can show that (25) holds in general, for all differentiable $f \in SG(\mathbb{R})$ with $f' \in SG(\mathbb{R})$ and all $\phi \in \mathcal{S}$.

Let us now rewrite (25) using the notation of (22). It becomes

$$L_{f'}(\phi) = -L_f(\phi'). \tag{26}$$

If we think of $L_{f'}$ as being, in some sense a "derivative" of L_f then this formula suggests how to define the derivative of a general tempered distribution:

Definition 3.2. *For each $L \in \mathcal{S}'$ let DL be the map from $\mathcal{S}$ to $\mathbb{C}$ defined by*

$$DL(\phi) = -L(\phi') \text{ for all } \phi \in \mathcal{S}.$$

*DL is called the **derivative** or the **distributional derivative** of L.*

We need a few moments to check that DL is also in $\mathcal{S}'$. Since L is linear, and $(\alpha\phi+\beta\psi)' = \alpha\phi' + \beta\psi'$ we immediately obtain that DL is linear. The second continuity condition is also easily verified: If $\{\phi_k\}_{k\in\mathbb{N}}$ is a sequence in $\mathcal{S}$ satisfying $\lim_{k\to\infty} C_{m,n}(\phi_k) = 0$ for all non negative integers m and n, then the sequence $\{\phi_k'\}_{k\in\mathbb{N}}$ has the same properties, simply because $C_{m,n}(\phi_k') = C_{m,n+1}(\phi_k)$. So $\lim_{k\to\infty} DL(\phi_k) = -\lim_{k\to\infty} L(\phi_k') = 0$.

We can now extend this definition to define the n^{th} order derivative of L to be the map D^nL given by

$$D^nL(\phi) = (-1)^n L(\phi^{(n)}) \text{ for all } \phi \in \mathcal{S}.$$

It is clear that $D^nL = D(D^{n-1}L)$ and that all these derivatives are elements of $\mathcal{S}'$.

The formula (26) tells us that whenever f is a differentiable function in $SG(\mathbb{R})$ such that f' is also in $SG(\mathbb{R})$ then $DL_f = L_{f'}$. We also see that $D\delta_0$ is given by $D\delta_0(\phi) = -\phi'(0)$. More generally, for each $a \in \mathbb{R}$ and $n \in \mathbb{N}$, the n^{th} derivative $D^n\delta_a$ is defined by $D^n\delta_a(\phi) = (-1)^n\phi^{(n)}(a)$ for all $\phi \in \mathcal{S}$. As our last example, consider the Heaviside function H already mentioned above and defined by $H(t) = \begin{cases} 0 & , \quad t<0 \\ 1 & , \quad t>0 \end{cases}$. It corresponds to the tempered distribution L_H defined by $L_H(\phi) = \int_0^\infty \phi(x)dx$ for all $\phi \in \mathcal{S}$. So its derivative, DL_H is defined by $DL_H(\phi) = -L_H(\phi') = -\int_0^\infty \phi'(x)dx = -(\lim_{r\to\infty}\phi(r) - \phi(0))$. Since the limit is 0 we obtain $DL_H(\phi) = \phi(0)$ for all $\phi \in \mathcal{S}$, i.e. we have now shown in an exact way that $DL_H = \delta_0$, the derivative of (the distribution corresponding to) H really is the delta "function".

I said in the heading to this section that $\mathcal{S}'$ is a space. Indeed it is a vector space. Given any L and M in $\mathcal{S}'$ and any complex numbers α and β we define the map $\alpha L + \beta M$ from $\mathcal{S}$ to $\mathbb{C}$ in the obvious way, i.e. $(\alpha L + \beta M)(\phi) = \alpha L(\phi) + \beta M(\phi)$ for each $\phi \in \mathcal{S}$. It is easy to check that $\alpha L + \beta M$, defined in this way, is also an element of $\mathcal{S}'$. The other conditions needed to show that $\mathcal{S}'$ is a vector space are easily verified[7]. For

[7] Cf. the general remarks about complex vector spaces in Section 7.

some remarks about the choice of the notation $\mathcal{S}'$ for the space of tempered distributions see Subsection 5.5 of the appendix.

Now we shall define the Fourier transforms of tempered distributions. The idea for doing this is in a similar spirit to what was done to define derivatives. First we find a suitable formula which is satisfied by Fourier transforms of functions. Then we "translate" this formula into something which also makes sense when we have distributions instead of functions.

Let us recall a formula for Fourier transforms of ***functions*** which is perhaps reminiscent of the generalized Plancherel formula. Before stating it, we should emphasize that, although it is "traditional" to use the variable x for a function f and the variable ω (or sometimes ξ) for its Fourier transform $\widehat{f}$, we have to be ready to change the names of these variables whenever necessary. For example, instead of saying that the Fourier transform of $e^{-|x|}$ is $\frac{1}{\pi(\omega^2+1)}$ it is equivalent, and maybe more precise to say: If $f:\mathbb{R}\to\mathbb{R}$ is the function defined by $f(t)=e^{-|t|}$ for all $t\in\mathbb{R}$, then the Fourier transform of f is the function $\widehat{f}:\mathbb{R}\to\mathbb{R}$ defined by $\widehat{f}(t)=\frac{1}{\pi(t^2+1)}$ for all $t\in\mathbb{R}$. The choice of the particular letters (here we chose t for both f and $\widehat{f}$) in the process of defining the function and its transform is completely unimportant. We could equally well choose x or ω, or any other letter, in each case. In particular, in the next result we will be writing $\widehat{f}(x)$ instead of the more traditional $\widehat{f}(\omega)$. But the meaning should be clear.

Lemma 3.3. *Suppose that f and ϕ are functions in $G(\mathbb{R})$. Then the product functions $f(x)\widehat{\phi}(x)$ and $\widehat{f}(x)\phi(x)$ are also in $G(\mathbb{R})$ and*

$$\int_{-\infty}^{\infty}\widehat{f}(x)\phi(x)dx=\int_{-\infty}^{\infty}f(x)\widehat{\phi}(x)dx. \tag{27}$$

Proof. By (1) $\widehat{f}$ and $\widehat{\phi}$ are bounded and continuous. This immediately implies that $f\widehat{\phi}$ and $\widehat{f}\phi$ are both in $G(\mathbb{R})$ and so both of the integrals in (27) exist. To prove that these integrals are equal we shall use Fubini's theorem[8] which enables us, ***provided that certain conditions are fulfilled,*** to change the order of integration in repeated integrals on $(-\infty,\infty)$. For details about this theorem see, for example, the web document,

http://www.math.technion.ac.il/˜mcwikel/FUBINI.PS

Let us consider the function $G(x,y)=f(x)\phi(y)e^{-iyx}$ on $\mathbb{R}^2$:

(i) $G(x,y)$ is "piecewise continuous" in a certain sense, (a product of piecewise continuous functions of one variable with a continuous function of two variables, as specified in equation (1) of the web document just referred to).

(ii) The limit $\lim_{N\to\infty}\int_{-N}^{N}\left(\int_{-N}^{N}|G(x,y)|dy\right)dx$ is finite. This is because $\int_{-N}^{N}\left(\int_{-N}^{N}|G(x,y)|dy\right)dx$ increases with N and it is bounded. It is bounded because

$$\int_{-N}^{N}\left(\int_{-N}^{N}|G(x,y)|dy\right)dx=\int_{-N}^{N}\left(\int_{-N}^{N}|f(x)\phi(y)|dy\right)dx$$

$$=\int_{-N}^{N}|f(x)|dx\cdot\int_{-N}^{N}|\phi(y)|\,dy\le\int_{-\infty}^{\infty}|f(x)|dx\cdot\int_{-\infty}^{\infty}|\phi(y)|\,dy$$

and this last expression is a product of two finite quantities which do not depend on N.

In view of the properties (i) and (ii) we can apply Fubini's theorem to obtain that

$$\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}G(x,y)dx\right)dy=\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}G(x,y)dy\right)dx$$

which is the same as

[8] This is perhaps the only result which you have not explicitly met before. But it is needed to justify some of the basic results for the Fourier transform which you presumably know (the inversion theorem, Plancherel formula and convolution property). It is used at least implicitly in proofs on pages 110, 115 and 118 of [PZ].

$$\int_{-\infty}^{\infty}\phi(y)\left(\int_{-\infty}^{\infty}f(x)e^{-iyx}dx\right)dy=\int_{-\infty}^{\infty}f(x)\left(\int_{-\infty}^{\infty}\phi(y)e^{-iyx}dy\right)dx.$$

If we divide both sides by 2π this is exactly (27). ■

It is the formula (27) which will tell us how to define the Fourier transform of elements of $\mathcal{S}'$: If ϕ happens to be a function in $\mathcal{S}$ then this formula can be rewritten as:

$$L_{\widehat{f}}(\phi)=L_f(\widehat{\phi}). \tag{28}$$

It seems reasonable to think of $L_{\widehat{f}}$ as the Fourier transform of L_f. So, to go one step further, if L is some other element of $\mathcal{S}'$ which is not generated by a function in $G(\mathbb{R})$, we can just copy (28) and define the Fourier transform of L to be a new map from $\mathcal{S}$ to $\mathbb{C}$, which we will denote by $\widehat{L}$ and which acts according to the formula

$$\widehat{L}(\phi)=L(\widehat{\phi})\text{ for all }\phi\in\mathcal{S}. \tag{29}$$

Here of course we need to know that $\widehat{\phi}\in\mathcal{S}$, but we have already shown that, as property 2.8. We now want to check that $\widehat{L}\in\mathcal{S}'$. First, since the Fourier transform of functions is linear, and L is linear we easily deduce that $\widehat{L}$ is also linear. Here is the proof:

$$\widehat{L}(\alpha\phi+\beta\psi)=L\left(\widehat{\alpha\phi+\beta\psi}\right)=L\left(\alpha\widehat{\phi}+\beta\widehat{\psi}\right)=\alpha L(\widehat{\phi})+\beta L(\widehat{\psi})=\alpha\widehat{L}(\phi)+\beta\widehat{L}(\psi).$$

So now it remains to show that $\widehat{L}$ satisfies the second "continuity" property of Definition 3.1. This is an immediate consequence of the following result:

Lemma 3.4. *Let $\{\phi_k\}_{k\in\mathbb{N}}$ be a sequence of functions in $\mathcal{S}$ which satisfies $\lim_{k\to\infty}C_{m,n}(\phi_k)=0$ for all non negative integers m and n. Then $\lim_{k\to\infty}C_{m,n}(\widehat{\phi_k})=0$ for all non negative integers m and n.*

The proof of Lemma 3.4, which will be deferred to an appendix (Subsection 5.6), uses the same ideas as were used to prove property 2.8 of $\mathcal{S}$, (in particular (1), (17) and (21)) with a more careful writing down of estimates.

Now let us calculate the Fourier transform of some particular distributions. For example, if $L=\delta_0$, then $\widehat{\delta_0}$ must satisfy

$$\widehat{\delta_0}(\phi)=\delta_0(\widehat{\phi})=\widehat{\phi}(0)=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-i0x}\phi(x)dx=\int_{-\infty}^{\infty}\frac{1}{2\pi}\phi(x)dx.$$

for all $\phi\in\mathcal{S}$. So we see that $\widehat{\delta_0}(\phi)=L_g(\phi)$ for all $\phi\in\mathcal{S}$, where g is the constant function $g(x)=\frac{1}{2\pi}$. This is the precise version of what we guessed to be true (cf. (2)) at the beginning of these notes.

More generally, suppose that $L=D^n\delta_a$ for some integer $n\geq 0$ and some $a\in\mathbb{R}$. Then, for every $\phi\in\mathcal{S}$, we see that $\widehat{L}=\widehat{D^n\delta_a}$ must satisfy

$$\widehat{D^n\delta_a}(\phi)=D^n\delta_a(\widehat{\phi})=(-1)^n\frac{d^n}{dx^n}\widehat{\phi}(a).$$

By (17) this equals $(-1)^n\cdot\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-iax}(-ix)^n\phi(x)dx=\int_{-\infty}^{\infty}\frac{1}{2\pi}e^{-iax}(ix)^n\phi(x)dx$. This means that $\widehat{D^n\delta_a}$ is a function. More precisely, it is the distribution L_f corresponding to the function $f(x)=\frac{1}{2\pi}e^{-iax}(ix)^n$.

Next we shall calculate the Fourier transform of x^n, i.e. of the distribution L_g corresponding to the function $g(x)=x^n$ for some integer $n\geq 0$. Perhaps you can already guess the answer from the previous example, if you suppose that there is a connection between Fourier transforms and inverse Fourier transforms of tempered distributions.

For every $\phi\in\mathcal{S}$, we have, using (29), (21) and then (7), that

$$\begin{aligned}\widehat{L_g}(\phi)&=L_g(\widehat{\phi})=\int_{-\infty}^{\infty}g(x)\widehat{\phi}(x)dx=\int_{-\infty}^{\infty}x^n\widehat{\phi}(x)dx=(-i)^n\int_{-\infty}^{\infty}(ix)^n\widehat{\phi}(x)dx\\&=(-i)^n\int_{-\infty}^{\infty}\widehat{\phi^{(n)}}(x)dx=(-i)^n\int_{-\infty}^{\infty}e^{i0x}\widehat{\phi^{(n)}}(x)dx=(-i)^n\phi^{(n)}(0)\\&=i^nD^n\delta_0(\phi).\end{aligned}$$

Since this is true for all $\phi \in \mathcal{S}$, we have just shown that $\widehat{L_{x^n}} = i^n D^n \delta_0$.

Finally let us extend this last example by linearity to show that the distributional Fourier transform of any polynomial $p(x) = \sum_{k=0}^{n} a_k x^k$ is given by $\sum_{k=0}^{n} a_k i^k D^k \delta_0$. An abbreviated way of writing this might be "$\widehat{p} = p(iD)\delta_0$". I enclosed this formula in quotation marks because both sides have to be interpreted carefully.

4. Some concluding remarks.

These few pages, as we already said, can only give a very quick introduction to tempered distributions. We shall conclude these notes by saying something about notation, and then considering and offering some quick and partial answers to two "natural" questions, an abstract one and then a slightly more "practical" one.

0. About notation.

In many books you will see that people write integrals with delta functions as if they were ordinary integrals, i.e., instead of the precise formula (in fact definition) that $\delta_0(\phi) = \phi(0)$ people like to write $\int_{-\infty}^{\infty} \delta_0(x)\phi(x)dx = \phi(0)$. More generally, if L is some general distribution in $\mathcal{S}'$ then sometimes people like to pretend that it is somehow like an ordinary function, and so, instead of the notation $L(\phi)$ which we have used here, they use the "integral" $\int_{-\infty}^{\infty} L(x)\phi(x)dx$ to denote the value of L when it acts on the function ϕ. This is fine and sometimes even useful, provided you remember that it is *only* notation, that it is *not* a real integral, and that in general, for individual values of x the symbol $L(x)$ may be *completely meaningless*. If you do not remember these things then you can easily get to all sorts of impossible and illogical conclusions, as we saw for example at the beginning of these notes.

Here is one example of how this "integral" notation can be used. Suppose L is a distribution and ϕ is a test function and c is a real constant. What is the exact meaning of the "integral" $\int_{-\infty}^{\infty} L(x+c)\phi(x)dx$? In other words, what is the distribution $L(x+c)$? Well, we could guess that we would like to have $\int_{-\infty}^{\infty} L(x+c)\phi(x)dx = \int_{-\infty}^{\infty} L(x)\phi(x-c)dx$, even though both of these things are not really integrals. But now we see that a logical interpretation for both of them is to say that, for each test function ϕ, they equal $L(\psi)$ where ψ is the test function defined by $\psi(x) = \phi(x-c)$ for all $x \in \mathbb{R}$.

1. How complicated and nasty can tempered distributions be?

Well they cannot be too terrible. It turns out that every tempered distribution can be constructed from a finite collection of continuous functions in a finite number of steps involving distributional differentiation. So we could say that tempered distributions are *relatively* simple objects. Here is a precise version of this result, the so called "structure theorem". (Cf. [Sc] pp. 239–240 for a slightly different but equivalent formulation. In future versions of these notes we will give additional references for this result.)

Given any tempered distribution L, there exists a finite collection of continuous functions $f_m : \mathbb{R} \to \mathbb{R}$, $m = 1, 2, ..., n$ which each satisfy $|f_m(x)| \le C(1+|x|)^N$ for some constant C and some integer N and all $x \in \mathbb{R}$ and such that L is given by

$$L = \sum_{m=1}^{n} D^m L_{f_m}.$$

This formula is of course the same as the condition

$$L(\phi) = \sum_{m=1}^{n} (-1)^m \int_{-\infty}^{\infty} f_m(x) \frac{d^m}{dx^m}\phi(x)dx \text{ for all } \phi \in \mathcal{S}.$$

2. What sort of things can distributions be used for?

I answer this as a pure mathematician. But I hope that, with the help of colleagues working in other fields, future versions of these notes will also mention other applications, such as those encountered in the courses which electrical engineering students take after they complete our basic course about Fourier transforms.

One important use of distributions is in the study of differential equations, including partial differential equations. In these notes we only considered functions of one variable. If we wish to consider partial differential equations we have to deal with functions of several variables, and their analogous generalized functions. There is a natural extension to $\mathbb{R}^n$ of the Fourier transform, the class $\mathcal{S}$, and so also the class $\mathcal{S}'$. Sometimes when it is not at all clear that a differential equation has a solution, i.e. a function satisfying the equation, it turns out that there are ways to show that there is a distribution which satisfies the same equation. Once it is known that there is a distributional solution, this can sometimes be the first step towards showing that there is also a solution in the original sense of the word. See [Ru], especially Chapter 8, for some examples of applications of distributions to partial differential equations.

5. Appendices.

5.1. **Functions in $\mathcal{S}$ which vanish outside a given interval.** The first and main step for constructing such functions is to consider the function $v : \mathbb{R} \to \mathbb{R}$ given by

$$v(x) = \begin{cases} e^{-1/x^2} \text{ if } x > 0 \\ 0 \text{ if } \leq 0 \end{cases}.$$

We will show that v is infinitely differentiable on $\mathbb{R}$, i.e. the n^{th} derivative $v^{(n)}(x)$ of v exists for each $n \in \mathbb{N}$ and for all $x \in \mathbb{R}$. For all $x > 0$ we have $v'(x) = \frac{2}{x^3}e^{-1/x^2}$ and $v''(x) = -\frac{6}{x^4}e^{-\frac{1}{x^2}} + \frac{4}{x^6}e^{-\frac{1}{x^2}}$. Both of these derivatives are finite sums of functions of the form $\frac{C}{x^m}e^{-1/x^2}$ where C and $m \in \mathbb{N}$ are constants. We do not need an explicit formula for $v^{(n)}$, but, by induction, we can continue and show that, for all $x > 0$ and each integer $n \geq 0$, $v^{(n)}(x)$ is a finite sum of functions of the form $\frac{C}{x^m}e^{-1/x^2}$. Since $\lim_{x \to 0+} \frac{1}{x^m}e^{-1/x^2} = \lim_{t \to \infty} t^m e^{-t^2} = 0$ for each $m \geq 0$, we see that $\lim_{x \to 0+} v^{(n)}(x) = 0$ and also that

$$\lim_{x \to 0+} \frac{1}{x} v^{(n)}(x) = 0. \tag{30}$$

In particular, for $n = 1$, this second limit shows that $\lim_{h \to 0+} \frac{v(h)-v(0)}{h} = 0$. Obviously also $\lim_{h \to 0-} \frac{v(h)-v(0)}{h} = 0$. So $v'(0)$ exists and equals 0. Obviously $v'(x)$ also exists for all $x \neq 0$.

Now let us use induction. If we know that $v^{(n)}(x)$ exists for all $x \in \mathbb{R}$ and also that $v^{(n)}(0) = 0$, then (30) tells us that $\lim_{h \to 0+} \frac{v^{(n)}(h)-v^{(n)}(0)}{h} = 0$ and, much as before, we deduce that $v^{(n+1)}(0) = 0$ and $v^{(n+1)}(x)$ exists for all $x \in \mathbb{R}$.

Now, given any bounded interval $[a, b]$, let $\phi(x) = v(x-a)v(b-x)$ for all $x \in \mathbb{R}$. Since $v(x-a)$ and $v(b-x)$ are both infinitely differentiable, so is $\phi(x)$ and obviously $\phi(x) = 0$ for all $x \leq a$ and all $x \geq b$, and $\phi(x) > 0$ for all $x \in (a, b)$. It is obvious that $\phi \in \mathcal{S}$.

5.2. **Recovering f from $L_f(\phi)$.**

Theorem 5.1. *Let ϕ be a fixed even function in $\mathcal{S}$ such that $\int_{-\infty}^{\infty} \phi(x)dx = 1$. For each $k \in \mathbb{N}$ and $c \in \mathbb{R}$ define $\phi_{k,c}$ by $\phi_{k,c}(x) = k\phi(k(x-c))$. Then, for each $f \in SG(\mathbb{R})$,*

$$f(c) = \lim_{k \to \infty} \int_{-\infty}^{\infty} f(x)\phi_{k,c}(x)dx. \tag{31}$$

In other words, we can determine the value of $f(c)$ at each point $c \in \mathbb{R}$ if we know the values of $L_f(\phi_{k,c})$ for each $k \in \mathbb{N}$. There are of course many ways to choose ϕ. For example we can take $\phi(x) = \frac{e^{-x^2}}{\sqrt{\pi}}$ or we can choose $\phi(x) = cv(x+a)v(a-x)$ for some $a > 0$, i.e. one of the functions constructed in Subsection 5.1 multiplied by a suitable constant c.

Proof. Some parts of the proof of this theorem may perhaps remind you of the proof in [PZ] of the inverse Fourier theorem. The proof can be made rather simpler in the case where the function ϕ vanishes outside some interval, but we shall treat the general case. We first use the change of variables $t = k(x-c)$ to obtain that

$$\begin{aligned} \int_{-\infty}^{\infty} f(x)k\phi(k(x-c))dx &= \int_{-\infty}^{\infty} f\left(c + \frac{t}{k}\right)\phi(t)dt \\ &= \int_{-\infty}^{0} f\left(c + \frac{t}{k}\right)\phi(t)dt + \int_{0}^{\infty} f\left(c + \frac{t}{k}\right)\phi(t)dt. \end{aligned}$$

We shall show that

$$\lim_{k \to \infty} \int_{0}^{\infty} f\left(c + \frac{t}{k}\right)\phi(t)dt = \frac{1}{2}f(c+) \text{ and } \lim_{k \to \infty} \int_{-\infty}^{0} f\left(c + \frac{t}{k}\right)\phi(t)dt = \frac{1}{2}f(c-). \tag{32}$$

Since f satisfies condition (5) this will give (31). We will only give the proof of the first part of (32) since the proof of the second part is almost exactly the same. Since ϕ is even and $\int_{-\infty}^{\infty}\phi(x)dx = 1$, we have $\int_0^\infty \phi(t)dt = \frac{1}{2}$ and so the first part of (32) is equivalent to the formula

$$\lim_{k\to\infty}\int_0^\infty \left(f\left(c+\frac{t}{k}\right)-f\left(c+\right)\right)\phi(t)dt = 0. \tag{33}$$

Given $\epsilon > 0$, there exists $\delta = \delta(\epsilon) > 0$ such that $|f(c+s) - f(c+)| \le \epsilon$ for all $s \in (0,\delta)$. Bearing this in mind we write

$$\left|\int_0^\infty \left(f\left(c+\frac{t}{k}\right)-f\left(c+\right)\right)\phi(t)dt\right| \le \int_0^\infty \left|\left(f\left(c+\frac{t}{k}\right)-f\left(c+\right)\right)\phi(t)\right|dt$$

$$\le \int_0^{\delta k} \left|\left(f\left(c+\frac{t}{k}\right)-f\left(c+\right)\right)\phi(t)\right|dt + \int_{\delta k}^\infty \left|\left(f\left(c+\frac{t}{k}\right)-f\left(c+\right)\right)\phi(t)\right|dt.$$

Then the first of the two integrals on the right is dominated by

$$\epsilon\int_0^{\delta k}|\phi(t)|dt \le \epsilon\int_0^\infty |\phi(t)|dt \tag{34}$$

and the second integral is dominated by

$$\int_{\delta k}^\infty |f\left(c+\right)\phi(t)|\,dt + \int_{\delta k}^\infty \left|f\left(c+\frac{t}{k}\right)\phi(t)\right|dt. \tag{35}$$

Since ϕ is absolutely integrable, the right side in (34) is finite, and the first integral in (35) equals $|f\left(c+\right)|\int_{\delta k}^\infty |\phi(t)|\,dt$ and tends to 0 as k tends to ∞. To estimate the second integral we finally have to use the fact that f satisfies the estimates (10) for some positive constants C and N. We also make two more changes of variable, first $y = t/k$, then later $z = c + y$. This gives

$$\begin{aligned}
\int_{\delta k}^\infty \left|f\left(c+\frac{t}{k}\right)\phi(t)\right|dt &= k\int_\delta^\infty |f\left(c+y\right)\phi(ky)|\,dy \\
&= k\sum_{n=1}^\infty \int_{\delta n}^{\delta(n+1)} |f\left(c+y\right)\phi(ky)|\,dy \\
&\le k\sum_{n=1}^\infty \frac{C_{N+2,0}(\phi)}{(k\delta n)^{N+2}}\int_{\delta n}^{\delta(n+1)} |f\left(c+y\right)|\,dy \\
&= k\sum_{n=1}^\infty \frac{C_{N+2,0}(\phi)}{(k\delta n)^{N+2}}\int_{c+\delta n}^{c+\delta(n+1)} |f\left(z\right)|\,dz \\
&\le k\sum_{n=1}^\infty \frac{C_{N+2,0}(\phi)}{(k\delta n)^{N+2}}\int_{-|c|-\delta(n+1)}^{|c|+\delta(n+1)} |f\left(z\right)|\,dz \\
&\le k\sum_{n=1}^\infty \frac{C_{N+2,0}(\phi)}{|k\delta n|^{N+2}}C(1+|c|+\delta(n+1))^N \\
&= k^{-N-1}C_{N+2,0}(\phi)C\sum_{n=1}^\infty \frac{1}{(\delta n)^{N+2}}(1+|c|+\delta(n+1))^N.
\end{aligned}$$

Using easy estimates similar to those in (14) we see that the general term in the series in this last expression is dominated by a constant multiple of $\frac{1}{n^2}$. So the series converges and $\int_{\delta k}^\infty \left|f\left(c+\frac{t}{k}\right)\phi(t)\right|dt \le k^{-N-1}M$ where the constant M does not depend on k. Combining all these estimates we obtain that

$$\limsup_{k\to\infty}\left|\int_0^\infty \left(f\left(c+\frac{t}{k}\right)-f\left(c+\right)\right)\phi(t)dt\right| \le \epsilon\int_0^\infty |\phi(t)|dt.$$

Then, since ϵ can be chosen arbitrarily small, this implies that

$$\limsup_{k\to\infty}\left|\int_0^\infty \left(f\left(c+\frac{t}{k}\right)-f\left(c+\right)\right)\phi(t)dt\right|=0$$

which gives (33) which, together with an analogous result for $\int_{-\infty}^0 \left(f\left(c+\frac{t}{k}\right)-f\left(c-\right)\right)\phi(t)dt$ completes the proof of the theorem. ■

5.3. **The distribution δ_0 is not given by a function in $SG(\mathbb{R})$.** This follows easily from Theorem 5.1. Suppose, on the contrary, that there exists a function $f\in SG(\mathbb{R})$ such that $\delta_0 = L_f$. This means that $\psi(0)=\delta_0(\psi)=L_f(\psi)=\int_{-\infty}^\infty f(x)\psi(x)dx$ for all $\psi\in\mathcal{S}$. In particular, if we apply this to the functions $\psi(x)=\phi_{k,c}(x)=k\phi(k(x-c))$ introduced in Theorem 5.1 we obtain that $f(c)=\lim_{k\to\infty}\int_{-\infty}^\infty f(x)\phi_{k,c}(x)dx = \lim_{k\to\infty}\phi_{k,c}(0)=\lim_{k\to\infty}k\phi(-kc)$. For each $c\neq 0$, since $|\phi(-kc)|\le C_{1,0}(\phi)/|kc|$, this last limit is 0. So $f(c)=0$ for all $c\neq 0$. For $c=0$ we may also obtain $f(0)=0$ if we choose a function ϕ satisfying $\phi(0)=0$, or otherwise, if $\phi(0)$ is real, we will obtain that $f(0)$ is ∞ or $-\infty$ depending on the sign of $\phi(0)$. These conclusions contain several contradictions: First of all the value of $f(0)$ as given by the formula (31) with $c=0$ should not depend on any particular choice of the function $\phi\in\mathcal{S}$ which satisfies $\int_{-\infty}^\infty \phi(x)dx=1$. It is supposed to be the same for all such ϕ. Then, $f(0)$ should be finite if $f\in SG(\mathbb{R})$. Finally, if we are talking about integrals in the usual sense of the word, changing the value of f at 0 should not influence the value of $\int_{-\infty}^\infty f(x)\phi(x)dx$ and this integral should thus equal 0 for all choices of $\phi\in S$, i.e. it will not in general equal $\phi(0)$. These contradictions show that δ_0 cannot equal L_f for any $f\in SG(\mathbb{R})$.

5.4. **Integration by parts on an infinite interval.** Let us complete the proof of (25) for all differentiable $f\in SG(\mathbb{R})$ and $\phi\in\mathcal{S}$ such that f' is also in $SG(\mathbb{R})$. In view of property 2.5, the functions $f'\phi$ and $f\phi'$ are both absolutely integrable on $\mathbb{R}$ and therefore also on $[0,\infty)$. So the limits $\lim_{r\to+\infty}\int_0^r f'(x)\phi(x)dx$ and $\lim_{r\to+\infty}\int_0^r f(x)\phi'(x)dx$ both exist. By (24), $f(r)\phi(r)=f(0)\phi(0)+\int_0^r f'(x)\phi(x)dx+\int_0^r f(x)\phi'(x)dx$. So, letting r tend to ∞, we see that the limit $\lim_{r\to+\infty}f(r)\phi(r)$ exists. Now suppose that $\lim_{r\to+\infty}f(r)\phi(r)=c>0$. Then for some sufficiently large r we will have $f(x)c(x)\ge c/2$ for all $x\in[r,\infty)$. But this is impossible since, again by property 2.5, the function $f\phi$ is also absolutely integrable. Similarly we get a contradiction if $c<0$. It follows that $\lim_{r\to+\infty}f(r)\phi(r)=0$. Similarly we can show that $\lim_{r\to-\infty}f(r)\phi(r)=0$ and so (25) follows from (24).

(We don't have to worry about this here, but in fact, it can be shown that f' has to be continuous on $\mathbb{R}$. This is because of a certain property of functions which are derivatives of other functions. Suppose that g is a function which is the derivative of some other function. There are examples which show that g does not have to be continuous. But g can only be discontinuous in certain ways. In particular, g cannot have any simple jump discontinuities, i.e, there are no points x_0 where $g(x_0+)$ and $g(x_0-)$ both exist and are different from each other.)

5.5. **A reason for using the notation $\mathcal{S}'$. Dual spaces, and continuous linear functionals.** The use of the notation $\mathcal{S}'$ is consistent with notation which is often used in mathematics, in particular in the field called ***functional analysis***. Suppose that V is a vector space (of functions, or of some other objects) and we have defined what we mean by convergent sequences in V. In particular this means that we have a definition of what it means for a given sequence $\{v_k\}_{k\in\mathbb{N}}$ to converge to the zero vector in V. (Sometimes this definition of convergence is made with the help of some norm on V. Sometimes, as in the case of $V=\mathcal{S}$ we prefer a different kind of definition.) . Then it is a standard procedure to define a new space, which is called the ***dual space*** of V, and which is denoted by V', or sometimes by V^*. This space consists of all "continuous" linear maps $L:V\to\mathbb{C}$. (Or if V is a vector space over $\mathbb{R}$ we will consider linear maps $L:V\to\mathbb{R}$) Here "continuous" means that, for every sequence $\{v_k\}_{k\in\mathbb{N}}$ which converges to the zero element of V we must have $\lim_{k\to\infty}L(v_k)=0$. The maps L are often called ***continuous linear functionals on*** V. In this framework we can see that the tempered distributions are exactly the continuous linear functionals on $\mathcal{S}$ and the space of all these distributions is the dual space of $\mathcal{S}$, so it is appropriate to use the notation $\mathcal{S}'$ for this space.

5.6. **The proof of Lemma 3.4.** For all $\phi \in \mathcal{S}$ and all $\omega \in \mathbb{R}$ we have

$$\begin{aligned} \left|\widehat{\phi}(\omega)\right| &\leq \frac{1}{2\pi}\int_{-\infty}^{\infty}|\phi(x)|dx = \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{1}{1+x^2}|(1+x^2)\phi(x)|dx \\ &\leq \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{1}{1+x^2}\cdot\left(C_{0,0}(\phi)+C_{2,0}(\phi)\right)dx = \frac{1}{2\pi^2}\cdot\left(C_{0,0}(\phi)+C_{2,0}(\phi)\right). \end{aligned} \tag{36}$$

Given any sequence $\{\psi_k\}_{k\in\mathbb{N}}$ of elements in $\mathcal{S}$ and any non negative integers m and n we first apply (17) to obtain that $\frac{d^n}{d\omega^n}\widehat{\psi_k}(\omega)$ is the Fourier transform of $\xi_k(x) = (-ix)^n\psi_k(x)$ which is also a function in $\mathcal{S}$. Then, by (21), $(i\omega)^m\widehat{\xi_k}(\omega)$ is the Fourier transform of $\xi_k^{(m)}(x)$ which of course is also in $\mathcal{S}$. Now, if we choose $\phi = \xi_k^{(m)}$, then $\widehat{\phi}(\omega) = (i\omega)^m\widehat{\xi_k}(\omega) = (i\omega)^m\frac{d^n}{d\omega^n}\widehat{\psi_k}(\omega)$. If we substitute in (36) this gives

$$\left|\omega^m\frac{d^n}{d\omega^n}\widehat{\psi_k}(\omega)\right| \leq \frac{1}{2\pi^2}\cdot\left(C_{0,0}(\xi_k^{(m)})+C_{2,0}(\xi_k^{(m)})\right).$$

Taking the supremum (or maximum) in this inequality as ω ranges over $\mathbb{R}$ gives

$$C_{m,n}(\widehat{\psi_k}) \leq \frac{1}{2\pi^2}\cdot\left(C_{0,0}(\xi_k^{(m)})+C_{2,0}(\xi_k^{(m)})\right).$$

From this inequality it is clear that to complete the proof of Lemma 3.4 we have to show that the condition

$$\lim_{k\to\infty} C_{m,n}(\psi_k) = 0 \text{ for all non negative integers } m \text{ and } n, \tag{37}$$

implies that $\lim_{k\to\infty} C_{0,0}(\xi_k^{(m)})$ and $\lim_{k\to\infty} C_{2,0}(\xi_k^{(m)})$ are both 0 for all non negative integers m and n. By Leibniz' formula, we have that

$$\xi_k^{(m)}(x) = \frac{d^m}{dx^m}\left[(-ix)^n\psi_k(x)\right] = \sum_{j=0}^{m}\binom{m}{j}\frac{d^{m-j}}{dx^{m-j}}(-ix)^n\cdot\psi_k^{(m)}(x), \tag{38}$$

where, as usual, $\binom{m}{j} = \frac{m!}{(m-j)!j!}$. Each term $\frac{d^{m-j}}{dx^{m-j}}(-ix)^n$ is 0 if $m-j>n$. Otherwise it equals $(-i)^n\frac{n!}{(n-m+j)!}x^{n-m+j}$. So (38) implies that

$$\left|\xi_k^{(m)}(x)\right| \leq \sum_{j=\max\{0,m-n\}}^{m}\binom{m}{j}\cdot\frac{n!}{(n-m+j)!}\cdot\left|x^{n-m+j}\psi_k^{(m)}(x)\right|$$

and so, for any integer $\gamma \geq 0$, we have

$$\begin{aligned} \left|x^\gamma\xi_k^{(m)}(x)\right| &\leq \sum_{j=\max\{0,m-n\}}^{m}\binom{m}{j}\cdot\frac{n!}{(n-m+j)!}\cdot\left|x^{n-m+j+\gamma}\psi_k^{(m)}(x)\right| \\ &\leq \sum_{j=\max\{0,m-n\}}^{m}\binom{m}{j}\cdot\frac{n!}{(n-m+j)!}\cdot C_{n-m+j+\gamma}(\psi_k) \end{aligned}$$

and so

$$0 \leq C_{\gamma,0}\left(\xi_k^{(m)}\right) \leq \sum_{j=\max\{0,m-n\}}^{m}\binom{m}{j}\cdot\frac{n!}{(n-m+j)!}\cdot C_{n-m+j+\gamma}(\psi_k). \tag{39}$$

By (37) the right side of (39) tends to 0 as $k\to\infty$ for each fixed m, n and γ. So $\lim_{k\to\infty} C_{\gamma,0}\left(\xi_k^{(m)}\right) = 0$ for each γ, in particular for $\gamma = 0$ and $\gamma = 2$, which, as explained above, is exactly what we need to complete the proof of Lemma 3.4. ■

6. A list of some of the symbols used in these notes

(The meanings of some of the terminology used in the definitions of these symbols are recalled in Section 7.)

• $\mathbb{R}$, $\mathbb{C}$, $\mathbb{N}$ and $\mathbb{Z}$. As usual these are, respectively, the sets of all real numbers, all complex numbers, all positive integers, and all integers.

• χ_E. Let E be any subset of $\mathbb{R}$, (in many cases E will be an interval). Then the ***characteristic function*** of E, sometimes also called the ***indicator function*** of E, is the function χ_E defined by $\chi_E(x) = \begin{cases} 1 & \text{if } x \in E \\ 0 & \text{if } x \notin E. \end{cases}$.

• $G(\mathbb{R})$. As in [PZ], $G(\mathbb{R})$ denotes the set of all functions $f : \mathbb{R} \to \mathbb{C}$ which are piecewise continuous on each bounded subinterval of $\mathbb{R}$ and which are absolutely integrable on $\mathbb{R}$.

• $\widehat{f}$. The Fourier transform $\widehat{f}$ of a function f is defined by slightly different formulae in different books. Here, as in [PZ], we use the formula $\widehat{f}(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\omega x} f(x)dx$. Later $\widehat{f}$ is also defined when f is a tempered distribution.

• PC_{bj}. The class of functions which are "piecewise continuous with balanced jumps". See the definition given immediately after (5).

• $SG(\mathbb{R})$. The set of functions of "slow growth". See Definition 1.4.

• $\mathcal{S}$. The Schwartz class of very smooth and very rapidly decaying functions. See Definition 2.1.

• $\mathcal{S}'$. The space of tempered distributions. See Definition 3.1.

7. Some reminders about some relevant mathematical notions.

Some of these notions and definitions can be found in the book [PZ], and some are introduced in these notes.

• **Piecewise continuity.**

Let E be a bounded interval. A function $f : E \to \mathbb{C}$ is said to be ***piecewise continuous on*** E if it is continuous at every point of E, or, in the worst case, there are only finitely many points of E at which f is not continuous and has a simple jump discontinuity. More precisely, we require the one–sided limits from the left and right, $\lim_{x\to c-0} f(x)$ and $\lim_{x\to c+0} f(x)$ to exist at every interior point c of E and to equal each other for all but at most finitely many interior points of E. If E contains its left endpoint a, then we also require the existence of $\lim_{x\to a+0} f(x)$. Similary if E contains its right endpoint b, then we require the limit $\lim_{x\to b-0} f(x)$ to exist. (However $f(a)$ and $f(b)$ do not have to equal these limits.) We should also mention other notation used for one sided limits: $f(c-)$, or $f(c-0)$ or $\lim_{x\nearrow c} f(x)$ for limits from the left, and $f(c+)$, or $f(c+0)$ or $\lim_{x\searrow c} f(x)$ for limits from the right.

If E is an unbounded interval, for example if $E = \mathbb{R}$, then we have adopted the same convention as in some books (cf. [PZ] p. 96). I.e., we define piecewise continuity of a function f on E to mean that f is piecewise continuous on every bounded subinterval of E. So f can possibly have infinitely many jumps altogether, but each bounded interval can contain at most finitely many of those jumps.

• **Riemann integrals of complex functions on a bounded interval** $[a, b]$**.**

We will not repeat the standard definitions and properties of Riemann integrals of real functions on an interval $[a, b]$. We only want to recall that these extend naturally and quite easily to the case of complex functions $f : [a, b] \to \mathbb{C}$. One way of doing this is to write each complex valued function in the form $f = u + iv$ where $u : [a, b] \to \mathbb{R}$ and $v : [a, b] \to \mathbb{R}$ are the real valued functions which are the real and imaginary parts of f. Then, depending on your taste, it is either a definition or a theorem that f is Riemann integrable on $[a, b]$ if and only if u and v are both Riemann integrable on $[a, b]$ and that $\int_a^b f(x)dx = \int_a^b u(x)dx + i\int_a^b v(x)dx$. Straightforward (but boring) calculations then show that the standard formulæ $\int_a^b \alpha f(x)+\beta g(x)dx = \alpha\int_a^b f(x)dx+\beta\int_a^b g(x)dx$ and $\left|\int_a^b f(x)dx\right| \le \int_a^b |f(x)|\, dx$ hold for all integrable f and g and all constants α and β also when some or all of them are complex valued.

• **Generalized Riemann integrals, and absolute integrability, on unbounded intervals.**

Suppose that $f : [0, \infty) \to \mathbb{C}$ has the property that its restriction to the interval $[0, R]$ is integrable for each constant $R > 0$. Then we can play two games:

(i) If the limit $\lim_{R\to\infty}\int_0^R f(x)dx$ and is finite, then we say that f has a ***generalized Riemann integral on*** $[0,\infty)$ which we denote and define by $\int_0^\infty f(x)dx = \lim_{R\to\infty}\int_0^R f(x)dx$.

(ii) If the limit $\lim_{R\to\infty}\int_0^R |f(x)|dx$ is finite[9] then we say that f is ***absolutely integrable on*** $[0,\infty)$.

It is not difficult to prove that if f is absolutely integrable, then f itself also has a generalized Riemann integral. (This is similar to the proof that an absolutely convergent series is also convergent.) But (analogously to series) the reverse implication is untrue[10].

On the interval $(-\infty,0]$ there are analogous definitions of generalized Riemann integrals $\int_{-\infty}^0 f(x)dx = \lim_{R\to+\infty}\int_{-R}^0 f(x)dx$ and absolute integrability for functions $f:(-\infty,0]\to\mathbb{C}$ which are Riemann integrable on each interval $[-R,0]$ for each $R>0$.

Finally, if the function $f:\mathbb{R}\to\mathbb{C}$ is Riemann integrable on $[-R,R]$ for each $R>0$, we can play three games:

(i) We say that f has a ***generalized Riemann integral on*** $(-\infty,\infty)$ if the integrals $\int_0^\infty f(x)dx$ and $\int_{-\infty}^0 f(x)dx$ defined above both exist. Then we define and denote this integral by $\int_{-\infty}^\infty f(x)dx = \int_{-\infty}^0 f(x)dx + \int_0^\infty f(x)dx$.

(ii) We say that f is ***absolutely integrable on*** $(-\infty,\infty)$ if $|f|$ has a generalized Riemann integral on $(-\infty,\infty)$.

(iii) If the "symmetric" limit $\lim_{R\to+\infty}\int_{-R}^R f(x)dx$ exists and is finite then we call its value the ***"Cauchy principal value" integral of*** f ***on*** $(-\infty,\infty)$, and use the notation $P.V.\int_{-\infty}^\infty f(x)dx = \lim_{R\to+\infty}\int_{-R}^R f(x)dx$ for this special kind of integral.

From what we have already said above, it is clear that the absolute integrability of f on $(-\infty,\infty)$, (which is equivalent to absolute integrability on both $[0,\infty)$ and $(-\infty,0]$) implies the existence of $\int_{-\infty}^\infty f(x)dx$, but that this integral also exists for functions f which are not absolutely integrable. The existence of $\int_{-\infty}^\infty f(x)dx$ obviously implies the existence of $P.V.\int_{-\infty}^\infty f(x)dx$ and then these two integrals are equal. If $f(x)\geq 0$ then $\int_{-\infty}^\infty f(x)dx$ exists if and only if $P.V.\int_{-\infty}^\infty f(x)dx$ exists. But simple examples, e.g. $f(x)=x$ show that $P.V.\int_{-\infty}^\infty f(x)dx$ can sometimes exist when $\int_{-\infty}^\infty f(x)dx$ does not.

One reason why principal value integrals $P.V.\int_{-\infty}^\infty f(x)dx$ are important is because of the formula for inverting the Fourier transform. We know that, if f is sufficiently nice, then $\frac{f(x+)+f(x-)}{2} = \lim_{R\to\infty}\int_{-R}^R \widehat{f}(\omega)e^{i\omega x}d\omega$ and this last integral is of course just $P.V.\int_{-\infty}^\infty \widehat{f}(\omega)e^{i\omega x}d\omega$. (The function $f:\mathbb{R}\to\mathbb{R}$ defined by $f(x)=\frac{x}{|x|}e^{-x}$ for all $x\neq 0$ and $f(0)=0$ provides an example showing that in general we cannot replace $\lim_{R\to\infty}\int_{-R}^R$ by $\int_{-\infty}^\infty$ in the Fourier inversion formula.)

• **Vector spaces or linear spaces.** A ***vector space over the complex field*** is a set V of elements, which we often call ***vectors***, on which we have defined two operations. The first operation is often called "addition of vectors" and usually denoted by $+$ (even though this same sign $+$ can have many other meanings in other contexts). The second operation is called multiplication of vectors by "scalars", which in our case are complex numbers[11] and is usually denoted by simply writing the scalar in front of the vector, i.e. λv for some scalar λ and vector v. These two operations have to satisfy a number of conditions, associative commutative and distributive laws, existence of a zero element, etc. which we will not list here. You can see these nine conditions, for example, on page 6 of [PZ]. We sometimes simply say ***vector space*** or ***linear space*** or ***complex vector space*** instead of "vector space over the complex field".

In these notes we encounter the following five examples of sets which can each be shown to be (infinite dimensional) complex vector spaces: $G(\mathbb{R})$, PC_{bj}, $SR(\mathbb{R})$, $\mathcal{S}$ and $\mathcal{S}'$. And of course $\mathbb{C}$ itself is also obviously a one dimensional complex vector space with the usual operations of addition and multiplication of complex numbers.

[9] Since $\int_0^R |f(x)|dx$ is a non decreasing function of R this limit always exists. But it may be infinite.

[10] Consider for example the function f which equals $(-1)^n/n$ on the interval $[n-1,n)$ for all $n\in\mathbb{N}$.

[11] The set of complex numbers $\mathbb{C}$ is only one example of an algebraic object called a which we will not define here. Other examples of fields are the set $\mathbb{R}$ of real numbers and the set $\mathbb{Q}$ of rational numbers. Our definition is a special case of the more general notion of a vector space over a field, which makes sense for any choice of field.

The sets $G(\mathbb{R})$, PC_{bj}, $SR(\mathbb{R})$, $\mathcal{S}$ and $\mathcal{S}'$ have some common features. Each one of them is a set of some kind of complex valued functions f defined on some kind of set Γ and the operations of vector addition and multiplication by scalars are defined "pointwise". (For each one of the sets $G(\mathbb{R})$, PC_{bj}, $SR(\mathbb{R})$ and $\mathcal{S}$ the set Γ is simply $\mathbb{R}$. In the more exotic case of $\mathcal{S}'$ we have to take $\Gamma = \mathcal{S}$.) This means that six of the required nine properties on page 6 of [PZ] follow immediately from analogous properties of multiplication and addition of complex numbers.

Let me explain this in a bit more detail: Suppose that V is one of the above five sets and that f and g are two vectors in V. Then they are both functions on Γ, i.e. for each $\gamma \in \Gamma$ we have complex numbers $f(\gamma)$ and $g(\gamma)$ which are the values of f and g respectively at the "point" γ. We define the new vector $f+g$ to be the function defined by the formula $(f+g)(\gamma) = f(\gamma) + g(\gamma)$ for all $\gamma \in \Gamma$. Note that in this last formula the symbol $+$ has two different meanings. On the left side it means the operation of addition in V, and on the right side it means the operation of addition of complex numbers. Similarly, for each $\lambda \in \mathbb{C}$, we define λf to be the function defined by $(\lambda f)(\gamma) = \lambda f(\gamma)$ for each $\gamma \in \Gamma$. (Here again, note that the writing of λ next to the vector f and the writing of λ next to the number $f(\gamma)$ have two different meanings.) Now we want to check that the operations defined in this way satisfy the nine conditions on page 6 of [PZ]. First we should consider conditions 1 and 5 which state that, whenever f and g are both in V and $\lambda \in \mathbb{C}$, then $f+g$ and λf both have to be in V. To verify these we need to use some special properties, depending on the particular choice of V. E.g., we know that sums of continuous functions are continuous, and sums of integrable functions are integrable, sums of linear maps are linear, etc. etc. We might also use the triangle inequality for complex numbers to show that certain required inequalities hold. Then condition 3 requires the existence of a special so–called zero element $\mathbf{0}$ in V with the property that $f + \mathbf{0} = f$ for each f $f \in V$. In each case that we have to consider, our set V contains the zero function, i.e the function whose value is 0 (the complex number) for each $\gamma \in \Gamma$. This clearly has the required property.

The verification of the remaining six conditions are, as I said above, almost automatic. For example, condition 2 is the associative law, that $(f+g)+h = f+(g+h)$ for all f, g and h in V. In our case this is simply the corresponding associative law for addition of complex numbers applied to the numbers $f(\gamma)$, $g(\gamma)$ and $h(\gamma)$ for each $\gamma \in \Gamma$.

• **Separating classes** (or **total sets**). Let A and B be classes of functions on $\mathbb{R}$ such that the integral $\int_{-\infty}^{\infty} f(x)g(x)dx$ exists for each $f \in A$ and each $g \in B$. Suppose that whenever f_1 and f_2 are functions in A such that $\int_{-\infty}^{\infty} f_1(x)g(x)dx = \int_{-\infty}^{\infty} f_2(x)g(x)dx$ for all $g \in B$, it follows that $f_1 = f_2$. Then we say that B is a ***separating class*** for A. (Or we can also say that B is a ***total set*** for A.)

8. Some books for further reading.

Here are some books, most of which I happened to notice in our mathematics library. I have surely missed some other very good books, and will be glad to add them to this list if you tell me about them.

[B] Hans Bremerman, Distributions, complex variables and Fourier transforms, Addison-Wesley 1965.

This book includes a historical introduction with some motivations from physics. The author attempts to simplify his presentation by delaying the use of difficult results from functional analysis until they are absolutely necessary.

[E] Robert Edward Edwards, Functional Analysis, Theory and Applications. Holt Rinehart and Winston 1965.

This is definitely not a book for beginners. Chapter 5 includes a very detailed and thorough treatment of distributions, with many references and comments.

[Ha] Israel Halperin, Introduction to the theory of distributions. University of Toronto Press, 1952.

This has the advantage of being short. But it does not deal with tempered distributions or Fourier transforms.

[Ho] Kenneth B. Howell, Principles of Fourier analysis. Studies in Advanced Mathematics. Boca Raton, FL., 2001.

There are more than 700 pages in this book in which the author patiently and thoroughly explains many aspects of Fourier analysis. In Section 27 (pp. 435–462) there is a naive description of the delta "function" and Heaviside function without precise definitions, and a plausible guess as to what the "Fourier transforms" *of these "functions" and of various periodic functions might be. Later in Chapter IV (pp. 497–694) the author gives exact definitions and a detailed study of generalized functions and their Fourier transforms. His approach is different from Laurent Schwartz's (and ours here). Instead of the Schwartz class $\mathcal{S}$ he uses a different class $\mathcal{G}$ of so-called "Gaussian test functions". See pp. 509–510 for a brief discussion of some advantages of using $\mathcal{G}$ instead of $\mathcal{S}$.*

[N] Hanna Neumann, Schwartz Distributions, Notes in Pure Mathematics, Australian National University.

This book appeared some time before 1968. It is not in our library and I do not know if it is still available anywhere. But it is a good, comparatively short introduction to the subject, including tempered distributions and Fourier transforms, designed to be understood by students with no background in functional analysis.

[PZ] Allan Pinkus and Sami Zafrany, Fourier Series and Integral Transforms.

The Hebrew version of this book is available at:

http://www.math.technion.ac.il/~ fourier/heb-ps.html

There is also an English version.

This is our reference in these notes for basic facts about Fourier transforms. It does not deal systematically with distributions, though the delta "function" is mentioned briefly and dealt with intuitively.

[RY] J. Ian Richards and Heekyung K. Youn, Theory of distributions: a non-technical introduction, Cambridge University Press 1990.

This is probably an easier book for non specialists than many of the other ones listed here. Among other things it includes Lars Hörmander's very clever and short proof of the Fourier inversion theorem for functions in $\mathcal{S}$ (on page 54.)

[Ru] Walter Rudin, Functional Analysis, McGraw Hill, 1973. *The relevant material is in Chapters 6, 7 and 8. In particular Chapter 8 gives some applications to partial differential equations. Rudin has a particular gift for presenting mathematical material with very short and very elegant proofs. (But sometimes it is useful to look at longer more complicated proofs to enhance our understanding in other ways.) I thank Jonathan Charbit for drawing my attention to this reference. 46RUD in our library.*

[Sc] Laurent Schwartz, Théorie des distributions, Hermann, Paris, 1966.

This is the second edition of the book by the man who created this theory. Our library only has it in French. I have not yet located an English verion of it, but I suppose it must exist somewhere.

[St] Robert Strichartz, A Guide to Distribution Theory and Fourier Transforms, Studies in Advanced Mathematics, CRC Press, Boca Raton, Ann Arbor, London, Tokyo, 1994.

Written in an attractive friendly style, with the goal of being accessible to students who have not studied all the topics normally assumed in books on this subject (such as Lebesgue integration). Not all results are proved. It gets to the relation between distributions and Fourier transforms relatively quickly.

Department of Mathematics, Technion - Israel Institute of Technology, Haifa 32000, Israel
E-mail address: mcwikel@math.technion.ac.il