Generalised functions and distributions

Hannes Gustafsson
ABSTRACT

Distribution theory is a branch of mathematical analysis which emerged during the first half of the 20th century, mainly from research on partial differential equations. It concerns distributions – objects that generalize the notion of real- and complex valued function which traditionally occurs in analysis – and establishes generalizations of concepts in differential calculus and Fourier analysis to distributions. In this thesis, we present the basic foundations of distribution theory and show the extensions of differentiation and the Fourier transform to distributions. Lastly, we briefly discuss and exemplify the application of the theory to linear partial differential equations.
## Contents

1. Introduction  
   
2. Preliminaries: The Fourier Transform  
   
3. Distributions  
   
4. Partial Differential Equations  

References  

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Introduction</td>
<td>5</td>
</tr>
<tr>
<td>2. Preliminaries: The Fourier Transform</td>
<td>6</td>
</tr>
<tr>
<td>3. Distributions</td>
<td>33</td>
</tr>
<tr>
<td>4. Partial Differential Equations</td>
<td>62</td>
</tr>
<tr>
<td>References</td>
<td>74</td>
</tr>
</tbody>
</table>
Acknowledgements

I want to express my sincere gratitude to my supervisor Anders Israelsson for his continuous support, encouragement, enlightening and constructive advice and feedback, and not least for the considerable patience he had with me throughout this project.

Secondly, I want to thank my co-supervisor Wulfgang Staubach for suggesting and introducing the topic to me, for letting me work on this project, and for his encouragement and helpful directions with regards to literature.
1. Introduction

Distribution theory is a relatively recent and modern branch of mathematical analysis which gradually emerged over the first half of the 20th century until being definitely established in its modern form in the formulation due to Laurent Schwartz [1] - a work which earned Schwartz the Fields Medal in 1950.

The theory presents a generalization of complex valued functions on Euclidean spaces, their differential calculus and the Fourier transform, akin to how Lebesgue integration theory acts as a generalization of Riemann integration. Both the roots of the theory and some of its greatest impacts can be found in the development of the theory of linear partial differential equations. The concepts of weak solution and weak derivatives, which are less stringent versions of the ordinary concepts involving integration, began emerging in the studies on differential equations in the late 1800s and early 1900s, for instance in the works of Du-Bois Reymond, Poincare, Zaremba, Bochner, Weyl and others [2]. In the 1930s, for the purpose of studying partial differential equations Sobolev more thoroughly developed concepts of weak derivatives, weak solutions and Sobolev spaces that in some sense involved the functional analytic foundation of distribution theory and moreover introduced essentially the first notion of distributions, in what became Sobolev theory [2,3], which is of major importance to linear partial differential equations. During the same period, attempts at generalizing the Fourier transform was also made, and the Dirac delta function appeared and saw use in physics [2] - which all eventually would be encompassed by the theory of distributions.

However, it was ultimately Schwartz who developed the theory to an even further extent in an remarkably elegant fashion, which definitely made it into an extensive and coherently formulated theory enveloping and connecting prior achievements.

Subsequently, distribution theory plays a central role in the theory of linear partial differential equations, an application which has yielded great results, for instance through the work of L. Hörmander [4].

In this thesis, we present the basic fundamentals of distribution theory as formulated by L. Schwartz. In particular, we show the extension of differential calculus and
the Fourier transform to distributions. The outline is as follows. We begin by a preliminary review of the classical Fourier transform on functions. Thereafter, we introduce distributions and show their basic properties, how they act as generalized functions and show the extension of differentiation and the Fourier transform to them. Finally, we conclude with a brief section about the application of distribution theory and the Fourier transform to partial differential equations. We follow more or less closely the slightly less technical texts on the subject by Strichartz [5] and Richards & Youn [6], although the ambition is to be as stringent as this permits and we attempt to occasionally fill in appropriate details. However, for the most part we avoid application of functional analysis and measure theory, apart from using a few necessary results from Lebesgue integration theory.

2. Preliminaries: The Fourier Transform

We begin by reviewing some basic and hopefully familiar, but important, theory regarding the Fourier transform of functions. This will be important later as the concepts will be generalized to distributions. The question whether integrals are absolutely convergent or convergent occurs often in Fourier analysis, which is evident from the definition of the Fourier transform. Many problems regarding the Fourier transform is therefore, naturally, essentially about integrability. This is most easily and satisfactorily treated with Lebesgue integration theory, but since the author is not too well-versed in this theory at the time of writing this work, and since it is not the main focus, we shall not delve much into details regarding it. In general we will not explicitly bother distinguishing whether integrals appearing throughout this work are Lebesgue or Riemann integrals, and we will use methods pertaining to both as we see fit, although this is not of any significant concern since the Riemann and Lebesgue integrals are equal whenever both exists. For those that are familiar with Lebesgue integration theory, we note that we will implicitly assume that all functions occurring herein are measurable. Some central and powerful results from Lebesgue integration theory will be used and stated without proofs, mainly the Theorem of dominated convergence as well as Tonelli’s and Fubini’s theorems, and in such cases we will at most comment briefly about it. With this disclaimer out
of the way, we begin by defining the setting of the Fourier transform of functions: integrable functions and $L^p$-spaces.

**Definition 2.1 (L^p-norms, L^p-spaces, integrability).** For $1 \leq p < \infty$ and $\Omega \subseteq \mathbb{R}^n$, the $L^p$-norm $\|f\|_{L^p(\Omega)}$ of a function $f : \Omega \rightarrow \mathbb{C}$ is defined as

$$\|f\|_{L^p(\Omega)} := \left( \int_{\Omega} |f(x)|^p \, dx \right)^{\frac{1}{p}}$$

We write $f \in L^p(\Omega)$ if $\|f\|_{L^p(\Omega)} < \infty$. In the special cases we abbreviate $L^p = L^p(\mathbb{R}^n)$, and call functions $f \in L^1$ integrable and functions $f \in L^2$ square-integrable. We define the $L^\infty$-norm $\|f\|_{L^\infty}$ to be

$$\|f\|_{L^\infty} = \inf \{ a \geq 0 : |f(x)| \leq a \text{ for almost every } x \}$$

and write $f \in L^\infty$ if $\|f\|_{L^\infty} < \infty$.

**Remark 2.2.**

(I) One can verify that the $L^p$ spaces are vector spaces over $\mathbb{R}$ (or even $\mathbb{C}$) with the usual addition and scalar multiplication of functions, and that in each case the $L^p$-norm indeed defines a norm on $L^p$. Moreover, $L^p$, $1 \leq p < \infty$, is also complete in the $L^p$-norm.

(II) In the case of $L^\infty$, we use the measure theoretic term "for almost every $x$", which means for every $x$ except at most on a set of measure zero - an equivalent expression is "almost everywhere", whenever it is clear which variable is referred to. Essentially, this means that $f \in L^\infty$ are bounded except possibly on a set which does not affect the integral of $f$. One should also note that, strictly speaking, the elements of $L^p$-spaces are equivalence classes of functions that are equal almost everywhere; that is, "=" in $L^p$ really means equal almost everywhere.

**Definition 2.3 (The Fourier Transform).** For a function $f : \mathbb{R}^n \rightarrow \mathbb{C}$, $f \in L^1$, define the Fourier transform of $f$, denoted $\mathcal{F}f$ or $\hat{f}$, as

$$\mathcal{F}f(\xi) = \hat{f}(\xi) = \int_{\mathbb{R}^n} f(x) e^{ix \cdot \xi} \, dx ; \; \xi \in \mathbb{R}^n$$
For $f \in L^2$, we define it instead to be

$$Ff := \lim_{k \to \infty} Ff_k$$

where $\{f_k\}_{k=1}^{\infty} \subset L^1 \cap L^2$ is any sequence converging to $f$ in $L^2$.

We may note that the property of being integrable over $\mathbb{R}^n$ (i.e. being $L^1$) is enough to ensure the existence of the Fourier transform of a function, because $|e^{ix \cdot \xi}| = 1$ for all $\xi \in \mathbb{R}^n$ and thus $|\hat{f}| \leq \|f\|_{L^1} < \infty$ and this guarantees that the integral defining $\hat{f}$ exists (is well-defined) and finite for all $\xi \in \mathbb{R}^n$. The Fourier transform can be defined on $L^2$ in the way done above due to the fact that $L^1 \cap L^2$ is dense in $L^2$, another fact which proof is omitted here.

Before we proceed with Fourier analysis, we will for completeness state without proofs two fundamental tools borrowed from measure- and integration theory. Firstly Fubini’s theorem, which will be called upon numerous times in the following, and secondly Lebesgue’s dominated converge theorem. For proofs and more details about these results we refer to [7] or to a textbook on the subject such as [8].

**Theorem 2.4** (Fubini’s theorem). Suppose $f(x, y) \in L^1(\mathbb{R}^n \times \mathbb{R}^m)$. Define $f_{y_0}(x) := f(x, y_0) : \mathbb{R}^n \to \mathbb{R}$ and $f_{x_0}(y) := f(x_0, y) : \mathbb{R}^m \to \mathbb{R}$. Then

(i) $f_{y_0}(x) \in L^1(\mathbb{R}^n)$ and $f_{x_0}(y) \in L^1(\mathbb{R}^m)$ for almost every $y_0 \in \mathbb{R}^n$, $x_0 \in \mathbb{R}^m$.

(ii) $F(y) = \int_{\mathbb{R}^n} f_y(x) \, dx \in L^1(\mathbb{R}^m)$ and $G(x) = \int_{\mathbb{R}^m} f_x(y) \, dy \in L^1(\mathbb{R}^n)$.

(iii) $\int_{\mathbb{R}^n \times \mathbb{R}^m} f(x, y) \, dx \, dy = \int_{\mathbb{R}^m} \left( \int_{\mathbb{R}^n} f_y(x) \, dx \right) \, dy = \int_{\mathbb{R}^n} \left( \int_{\mathbb{R}^m} f_x(y) \, dy \right) \, dx$

**Remark 2.5.** The main, important conclusion for our purposes is (iii), since it dictates that one can change the order of integration and maintain equality; the other are merely formalities in this context. Tonelli’s theorem is closely related to Fubini’s theorem and we do not bother to state it in full, but it partly asserts that if the function $f(x, y)$ is non-negative but not necessarily in $L^1(\mathbb{R}^n \times \mathbb{R}^m)$, (iii) still
holds, but in the extended sense, so that all three integrals either is finite and equal, or all diverge to \(+\infty\).

**Theorem 2.6** (Lebesgue’s dominated convergence theorem). Suppose that \(\{f_k\}\) is a sequence of measurable functions on a measurable set \(E \subset \mathbb{R}^n\), \(f_k \to f\) as \(k \to \infty\) and there exists an integrable function \(g\) on \(E\) such that \(g(x) \geq |f_k(x)|\) for all \(x \in E\) and for all \(k\). Then

\[
\lim_{k \to \infty} \int_E f_k(x) \, dx = \int_E f(x) \, dx
\]

We now begin with an example by calculating a particular Fourier transform, which will be very useful later, namely the Fourier transform of a Gaussian function \(e^{-ax^2}\). Here we follow the discussion and calculation in [5], possibly including more details.

**Example 2.7** (Fourier Transform of \(e^{-ax^2}\)). Let \(a \geq 0\). First in \(\mathbb{R}\).

\[
\mathcal{F}(e^{-ax^2}) = \int_{-\infty}^{\infty} e^{-ax^2} e^{ix \xi} \, dx = \int_{-\infty}^{\infty} e^{-a(x^2 - \frac{ix \xi}{a})} \, dx
\]

By completing the square in the exponent, this becomes:

\[
(1) \quad \int_{-\infty}^{\infty} e^{-a(x^2 - \frac{ix \xi}{2a})} \, dx = \int_{-\infty}^{\infty} e^{-a(x - \frac{ix}{2a})^2} \frac{\xi^2}{4a} \, dx = e^{-\frac{\xi^2}{4a}} \int_{-\infty}^{\infty} e^{-a(x - \frac{ix}{2a})^2} \, dx
\]

To compute this integral, we consider the integral with the same integrand but counter-clockwise around a rectangle \(\Gamma_R\) in the \(\mathbb{C}\)-plane with vertices in \(-R, R, R + i\frac{\xi}{2a}\) and \(-R + i\frac{\xi}{2a}\). See Figure 1. We are interested in the limit \(R \to +\infty\). Since

\[\text{Figure 1. The rectangular contour } \Gamma_R \text{ in the complex plane, over which the integral is taken, shown in red.}\]
the integrand is analytic on and inside $\Gamma_R$, Cauchy’s theorem from complex analysis implies

$$0 = \int_{\Gamma_R} e^{-a(z-\frac{i\xi}{2a})^2} \, dz = \int_{-R}^{R} e^{-a(z-\frac{i\xi}{2a})^2} \, dz + \int_{R+i\frac{\xi}{2a}}^{R+\frac{\xi}{2a}} e^{-a(z-\frac{i\xi}{2a})^2} \, dz + \int_{R-i\frac{\xi}{2a}}^{-R-i\frac{\xi}{2a}} e^{-a(z-\frac{i\xi}{2a})^2} \, dz + \int_{-R-i\frac{\xi}{2a}}^{-R+i\frac{\xi}{2a}} e^{-a(z-\frac{i\xi}{2a})^2} \, dz$$

(2)

Now consider the integrals on the vertical sides of $\Gamma_R$, i.e. parallel to the imaginary axis. By parametrizing $z$ as $z(s) = R + i(s + 1)\frac{\xi}{2a}$, $s \in [-1,0]$ we obtain by applying the triangle inequality:

$$\left| \int_{R+i\frac{\xi}{2a}}^{R+\frac{\xi}{2a}} e^{-a(z-\frac{i\xi}{2a})^2} \, dz \right| = \left| \int_{-1}^{0} e^{-a(R+s\frac{\xi}{2a})^2} \frac{i\xi}{2a} \, ds \right|$$

$$\leq e^{-aR^2} \int_{-1}^{0} \left| e^{-i\pi aR} e^{s^2 \frac{\xi^2}{4a}} \frac{i\xi}{2a} \right| \, ds$$

$$= \left| \frac{\xi}{2a} \right| e^{-aR^2} \int_{-1}^{0} e^{s^2} \frac{\xi^2}{4a} \, ds$$

$$\leq \left| \frac{\xi}{2a} \right| e^{-aR^2} e^{\frac{\xi^2}{4a}}$$

(3)

The last integral in (3) is independent of $R$, and is finite since it is bounded above by $\sup_{-1 \leq s \leq 0} e^{s^2} = e^{\frac{\xi^2}{4a}}$. Clearly then the last expression in (3) approaches 0 as $R \to +\infty$ due to the factor $e^{-aR^2}$, proving that the vertical integral is 0 in the limit. Virtually the same argument can be used to show the same about $\int_{-R-i\frac{\xi}{2a}}^{-R+i\frac{\xi}{2a}} e^{-a(z-\frac{i\xi}{2a})^2} \, dz$. Thus by taking the limit $R \to \infty$ in (2) we obtain

$$\lim_{R \to \infty} \left( \int_{-R}^{R} e^{-a(z-\frac{i\xi}{2a})^2} \, dz + \int_{R+i\frac{\xi}{2a}}^{R+\frac{\xi}{2a}} e^{-a(z-\frac{i\xi}{2a})^2} \, dz \right) = 0$$

Equivalently,

$$\int_{-\infty}^{\infty} e^{-a(z-\frac{i\xi}{2a})^2} \, dz = \lim_{R \to \infty} \int_{-R-i\frac{\xi}{2a}}^{-R+i\frac{\xi}{2a}} e^{-a(z-\frac{i\xi}{2a})^2} \, dz$$

(4)

With the parametrization $z(s) = s + i\frac{\xi}{2a}$ we have

$$\int_{-R-i\frac{\xi}{2a}}^{-R+i\frac{\xi}{2a}} e^{-a(z-\frac{i\xi}{2a})^2} \, dz = \int_{-R}^{R} e^{-as^2} \, ds$$

(5)
Finally, combining (1), (4) and (5) we arrive at
\[
\mathcal{F}(e^{-ax^2})(\xi) = e^{-\frac{\xi^2}{4a}} \int_{-\infty}^{\infty} e^{-a(z - \frac{i\xi}{2a})^2} dz
\]
\[
e^{-\frac{\xi^2}{4a}} \lim_{R \to +\infty} \int_{-R+i\frac{\xi}{2a}}^{R+i\frac{\xi}{2a}} e^{-a(z - \frac{i\xi}{2a})^2} dz
\]
\[
e^{-\frac{\xi^2}{4a}} \lim_{R \to +\infty} \int_{-R}^{R} e^{-as^2} ds
\]
\[
= e^{-\frac{\xi^2}{4a}} \sqrt{\frac{\pi}{a}}
\]
In the last equality the fact that \( \int_{-\infty}^{\infty} e^{-as^2} ds = \sqrt{\frac{\pi}{a}} \) was used. Now we may move on to \( \mathbb{R}^n \) by noting that \( e^{-ax^2} = e^{-a(x_1^2 + x_2^2 + \ldots + x_n^2)} = e^{-ax_1^2} \ldots e^{-ax_n^2} \), and \( e^{ix \cdot \xi} = e^{ix_1 \xi_1} \ldots e^{ix_n \xi_n} \) so that the Fourier transform integral reduces to a product of one-dimensional Fourier transforms:
\[
\mathcal{F}e^{-ax^2}(\xi) = \int_{\mathbb{R}^n} e^{-ax^2} e^{ix \cdot \xi} dx
\]
\[
= \int_{\mathbb{R}^n} e^{-ax_1^2} \ldots e^{-ax_n^2} e^{ix_1 \xi_1} \ldots e^{ix_n \xi_n} dx
\]
\[
= \int_{\mathbb{R}^1} e^{-ax_1^2} e^{ix_1 \xi_1} dx_1 \ldots \int_{\mathbb{R}^1} e^{-ax_n^2} e^{ix_n \xi_n} dx_n
\]
\[
= e^{-\frac{\xi_1^2}{4a}} \sqrt{\frac{\pi}{a}} \ldots e^{-\frac{\xi_n^2}{4a}} \sqrt{\frac{\pi}{a}}
\]
\[
= e^{-\frac{\xi^2}{4a}} \left( \sqrt{\frac{\pi}{a}} \right)^n
\]
The Fourier transform of a Gaussian is thus also a Gaussian function.

An operation highly related to Fourier transforms is the convolution, introduced below. The Fourier transform and the convolution operation intermingle in very useful and convenient ways.

**Definition 2.8 (Convolution of Functions).** For functions \( f, g \in L^1 \) (or \( L^2 \)), define their convolution \( f \ast g \) to be
\[
f \ast g(x) := \int_{\mathbb{R}^n} f(x - y)g(y) dy
\]
given that the integral exists.

Here, a problem of integrability arises again. For instance, even if both \( f(x) \) and \( g(x) \) are absolutely integrable, the product \( f(x)g(x) \) may still fail to be. However, it
is rather easy to see that if for instance $f$ is $L^1$ and $g$ is bounded, then the product is $L^1$. With some measure theory, one can relax the condition on $g$ by requiring only that $g \in L^\infty$, which is to say that $g$ is bounded almost everywhere, so that $g$ is effectively bounded on all sets which impact the integral of $f(x-y)g(y)$. One can however go even further, as it turns out that $f, g \in L^1$ is enough to ensure that $f(x-y)g(y)$ is integrable in $y$ for almost every $x$, and thus the convolution is always defined on $L^1$, and moreover $f * g \in L^1$ and

$$\|f * g\|_{L^1} \leq \|f\|_{L^1} \|g\|_{L^1} < \infty$$

This is a consequence of Tonelli’s and Fubini’s theorems, although this is beyond the scope of this thesis. We direct interested readers to [8].

It is easier to see that the convolution is defined on $L^2$ and in that case always yields a bounded function, because this is a consequence of the Cauchy-Schwarz inequality and translation-reflection invariance of the $L^p$-norms. Indeed, if we let $f, g \in L^2$, and define $f_x(y) = f(x-y)$, then $\|f_x\|_{L^2} = \|f\|_{L^2}$ so $f_x \in L^2$, and by Cauchy-Schwarz

$$|f * g(x)| = \left| \int_{\mathbb{R}^n} f(x-y)g(y) \, dy \right| = |\langle f_x, g \rangle| \leq \|f\|_{L^2} \|g\|_{L^2} < \infty.$$

Thus the convolution operation is well-defined on $L^1$ and $L^2$, and additionally $f * g$ is bounded when $f, g \in L^2$ or if, say, $f \in L^1$ and $g \in L^\infty$. The convolution operation has the following properties.

**Proposition 2.9 (Properties of the convolution).** For functions $f, g, h \in L^1$,

(i) $f * g(x) = g * f(x)$

(ii) $f * (g * h)(x) = (f * g) * h(x)$

If additionally $f \in C^2$, $\frac{\partial f}{\partial x_j} \in L^1$ and $\frac{\partial^2 f}{\partial x_j^2}$ is bounded, then

(iii) $\frac{\partial}{\partial x_j} (f * g) = \left( \frac{\partial f}{\partial x_j} \right) * g$

**Proof.** (i) is easily shown with just a simple change of variables; (ii) can be shown by a change of variables and a change in the order of integrations. All of the variable substitutions are linear with all entries in the corresponding Jacobian matrices equal to either 0 or ±1, and such that there is only one non-zero term in the determinant sum. Thus, the modulus of the Jacobian determinant in each case is 1.
(i) \( f \ast g(x) = \int_{\mathbb{R}^n} f(x-y)g(y) \, dy = \begin{cases} \quad z = x - y \\ \quad y = z - x \end{cases} \)

\[
= \int_{\mathbb{R}^n} g(x-z)f(z) \, dz
= g \ast f(x)
\]

The change of variables required for (ii) can be justified by considering \( |f\ast(|g\ast|h)| \). From the assumptions and the discussion leading up to this proposition, \( |g\ast|h| \in L^1(\mathbb{R}^n) \), hence \( |f\ast(|g\ast|h)| \in L^1(\mathbb{R}^n) \), and the latter is finite for almost every \( x \in \mathbb{R}^n \). This implies, using Tonelli’s theorem, that \( H(y,z) = f(x-y)g(y-z)h(z) \in L^1(\mathbb{R}^n \times \mathbb{R}^n) \) w.r.t. \( (y,z) \) for almost every \( x \in \mathbb{R}^n \), so for all such \( x \) Fubini’s theorem then validates steps (\( \ast \)) and (\( \ast \)) in the below calculation.

(ii) \( f \ast (g \ast h)(x) = \int_{\mathbb{R}^n} f(x-y)g \ast h(y) \, dy \)

\[
= \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} f(x-y)g(y-z)h(z) \, dz \, dy
= \begin{cases} \quad \sigma = y-z \\ \quad z' = z \end{cases}
\rightarrow x-y = (x-z') - \sigma
\]

\[
= \int_{\mathbb{R}^n \times \mathbb{R}^n} f((x-z') - \sigma)g(\sigma)h(z') \, d\sigma \, dz'
= \begin{cases} \quad \sigma = y-z \\ \quad z' = z \end{cases}
\rightarrow x-y = (x-z') - \sigma
\]

\[
= \int_{\mathbb{R}^n} \left( \int_{\mathbb{R}^n} f((x-z') - \sigma)g(\sigma) \, d\sigma \right) h(z') \, dz'
= \int_{\mathbb{R}^n} f \ast g(x-z)h(z) \, dz
= (f \ast g) \ast h(x)
\]

To show the last statement, we use the fact that \( f \in C^2 \) implies that there is \( \theta_h \in [0,h] \) (or \( [h,0] \) if \( h < 0 \)), such that

\[
f(x-y + h\mathbf{e}_j) - f(x-y) = h \frac{\partial f}{\partial x_j}(x-y) + \frac{h^2}{2} \frac{\partial^2 f}{\partial x^2_j}(x-y + \theta_h\mathbf{e}_j)
\]

where \( \mathbf{e}_j \) is a unit vector in the \( x_j \) direction and \( \theta_h \in [0,h] \) (or \( [h,0] \) if \( h < 0 \)). We can then write

\[
\frac{\partial}{\partial x_j} \int_{\mathbb{R}^n} f(x-y)g(y) \, dy = \lim_{h \to 0} \frac{1}{h} \left( \int_{\mathbb{R}^n} (f(x-y + h\mathbf{e}_j) - f(x-y)) g(y) \, dy \right)
\]

\[
= \int_{\mathbb{R}^n} \frac{\partial f}{\partial x_j}(x-y)g(y) \, dy + \lim_{h \to 0} \frac{h}{2} \int_{\mathbb{R}^n} \frac{\partial^2 f}{\partial x^2_j}(x-y + \theta_h\mathbf{e}_j) g(y) \, dy
\]

(7)
Now from the assumption that $\frac{\partial^2 f}{\partial x_j^2}$ is bounded, there is some constant $M$ such that $\left| \frac{\partial^2 f}{\partial x_j^2}(x) \right| < M$ for all $x \in \mathbb{R}^n$. We then have

$$\lim_{h \to 0} \frac{h}{2} \int_{\mathbb{R}^n} \frac{\partial^2 f}{\partial x_j^2}(x - y + \theta_h e_j) g(y) \, dy \leq \lim_{h \to 0} \frac{h}{2} M \int_{\mathbb{R}^n} |g(y)| \, dy$$

(8)

$$= \lim_{h \to 0} \frac{h}{2} M \|g\|_{L^1} = 0$$

Then we see from (7) that

$$\frac{\partial}{\partial x_j}(f \ast g)(x) = \frac{\partial}{\partial x_j} \int_{\mathbb{R}^n} f(x - y)g(y) \, dy = \int_{\mathbb{R}^n} \frac{\partial f}{\partial x_j}(x - y)g(y) \, dy = \frac{\partial f}{\partial x_j} \ast g(x)$$

which is (iii).

Note that (i) above clearly also holds for $f, g \in L^2$ as well, by the same argument as above. For (ii) on the other hand, because it is not evident whether $f \ast g$ is necessarily $L^1$ or $L^2$ even if $f, g \in L^2$, it is unclear if $f \ast (g \ast h)$ is even well-defined in this case. However, we can see that if $f \in L^1$ and $g, h \in L^2$, then the integral defining $f \ast (g \ast h)$ is absolutely convergent since $g \ast h$ is bounded, and we may then apply the same argument as above to show (ii). By (i), we see that (ii) then holds for functions in $L^2$ if at least one of $f, g, h$ is integrable.

Next some elementary properties of the Fourier transform are stated, concerning its interaction with convolutions and differentiation. These make up a main cornerstone why the Fourier transform is highly useful when applied to linear partial differential equations, which will be exemplified later in the last section.
Proposition 2.10 (Properties of $\mathcal{F}$). If $f, g \in L^1(\mathbb{R}^n)$ the following holds:

(i) $\mathcal{F}(af + bg)(\xi) = a\mathcal{F}f(\xi) + b\mathcal{F}g(\xi)$ for constants $a, b \in \mathbb{C}$

(ii) $\mathcal{F}f$ is bounded and continuous on $\mathbb{R}^n$.

(iii) If $f \in C^1$, $\frac{\partial f}{\partial x_j} \in L^1(\mathbb{R}^n)$ and $f(x) \to 0$ when $|x_j| \to \infty$, then

$$\mathcal{F} \left( \frac{\partial f}{\partial x_j} \right)(\xi) = -i\xi_j \mathcal{F}f(\xi)$$

(iv) If $x_j f(x) \in L^1(\mathbb{R}^n)$ and $x_j^2 f(x) \in L^1(\mathbb{R}^n)$, then

$$\mathcal{F}(x_j f(x))(\xi) = -i \frac{\partial}{\partial \xi_j} \mathcal{F}f(\xi)$$

(v) $\mathcal{F}(f * g) = \mathcal{F}f \mathcal{F}g$

Proof.

(i) $\mathcal{F}(af + bg)(\xi) = \int_{\mathbb{R}^n} (af(x) + bg(x))e^{ix\cdot\xi} \, dx$

$$= a \int_{\mathbb{R}^n} f(x)e^{ix\cdot\xi} \, dx + b \int_{\mathbb{R}^n} g(x)e^{ix\cdot\xi} \, dx$$

$$= a\mathcal{F}f(\xi) + b\mathcal{F}g(\xi)$$

(ii) Boundedness was already shown before: $|\mathcal{F}f| \leq ||f||_{L^1}$. Continuity is a consequence of Theorem 2.6

(iii) Integration by parts in the $x_j$ variable gives:

$$\mathcal{F} \left( \frac{\partial f}{\partial x_j} \right)(\xi) = \int_{\mathbb{R}^n} \frac{\partial f}{\partial x_j}(x)e^{ix\cdot\xi} \, dx$$

$$= \int_{\mathbb{R}^{n-1}} \left[ f(x)e^{ix\cdot\xi} \right]_{x_j=\infty}^{x_j=-\infty} \, dx - \int_{\mathbb{R}^n} f(x)i\xi_j e^{ix\cdot\xi} \, dx$$

$$= 0 - i\xi_j \int_{\mathbb{R}^n} f(x)e^{ix\cdot\xi} \, dx$$

$$= -i\xi_j \mathcal{F}f(\xi)$$

The integral containing the boundary terms are zero due to the assumption that $f$ vanishes as $|x_j| \to \infty$. The integrability conditions makes sure that the integrals are finite, and enables application of Fubini’s theorem to iterate the integral and perform integration by parts in the $x_j$-integral.
(iv) Using that \( \frac{\partial}{\partial \xi_j}(e^{ix\cdot \xi}) = ix_j e^{ix\cdot \xi} \),

\[
\mathcal{F}(x_j f)(\xi) = \int_{\mathbb{R}^n} x_j f(x) e^{ix\cdot \xi} \, dx
\]

\[
= \int_{\mathbb{R}^n} f(x)(-i)(ix_j e^{ix\cdot \xi}) \, dx
\]

\[
= (-i) \int_{\mathbb{R}^n} \frac{\partial}{\partial \xi_j} (f(x) e^{ix\cdot \xi}) \, dx
\]

\[
\overset{(\ast)}{=} -i \frac{\partial}{\partial \xi_j} \int_{\mathbb{R}^n} f(x) e^{ix\cdot \xi} \, dx
\]

\[
= -i \frac{\partial}{\partial \xi_j} \mathcal{F}f(\xi)
\]

In (\ast), we interchange the order of the partial derivative with the integral. We can do this due to a very similar argument as was used to show (iii) in Proposition 2.9, by using that \( e^{ix\cdot \xi} \in C^2, \quad |e^{ix\cdot \xi}| \leq 1 \), and the assumption that \( x_j^2 f(x) \in L^1 \). The integral that arises in this case, corresponding to the integral in (8), will have an integrand which is a product of an integrable function and a bounded function, so the integral is finite. From this it is clear that

\[
\lim_{h \to 0} \frac{h}{2} \int_{\mathbb{R}^n} (ix_j)^2 f(x) e^{ix\cdot \xi} \, dx = 0
\]

and thus we see that the same argument works.

(v) \( \mathcal{F}(f \ast g)(\xi) = \int_{\mathbb{R}^n} \left( \int_{\mathbb{R}^n} f(x-y)g(y) \, dy \right) e^{ix\cdot \xi} \, dx \)

\[
= \int_{\mathbb{R}^n} \left( \int_{\mathbb{R}^n} f(x-y)e^{i(x-y)\cdot \xi} g(y)e^{iy\cdot \xi} \, dy \right) \, dx
\]

\[
= \int_{\mathbb{R}^n \times \mathbb{R}^n} g(y)e^{iy\cdot \xi} f(x-y)e^{i(x-y)\cdot \xi} \, dx \, dy = \left\{ \begin{array}{c} z = x - y \\ y' = y \end{array} \right\}
\]

\[
= \int_{\mathbb{R}^n \times \mathbb{R}^n} g(y')e^{iy'\cdot \xi} f(z)e^{iz\cdot \xi} \, dz \, dy'
\]

\[
= \mathcal{F}g(\xi) \mathcal{F}f(\xi)
\]

Clearly the linear change of variables is such that the Jacobian matrix \( J \) satisfies \( |\det J| = 1 \). We may again freely go back and forth between iterated
integrals in any order single integrals over $\mathbb{R}^n \times \mathbb{R}^n$ due to the integrability of $f * g$ and Fubini’s theorem.

\[ \square \]

**Remark 2.11.** Both (iii) and (iv) may with the use of induction and linearity (i) be generalized to polynomials $p(x)$ in $x = (x_1, \ldots, x_n)$ and polynomial expressions $p \left( \frac{\partial}{\partial x} \right)$ of $\frac{\partial}{\partial x} = \left( \frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_n} \right)$ (or alternatively, to $\frac{\partial^\alpha}{\partial x^\alpha}$ for any multi-index $\alpha$), yielding

\[ \text{(iii)'}\quad \mathcal{F} \left( p \left( \frac{\partial}{\partial x} \right) f \right) = p(-i\xi) \mathcal{F} f (\xi) \]

\[ \text{(iv)'}\quad \mathcal{F} (p(x) f) = p \left( -i \frac{\partial}{\partial \xi} \right) \mathcal{F} f (\xi) \]

Of course, this demands similar but appropriately stronger assumptions on $f$ and its partial derivatives up to some sufficiently high order.

A significant portion of the usefulness of the Fourier Transform when applied to problems, relies first of all on the properties that derivatives (linear differential operators) are transformed into (multiplication with) polynomials. Thus it for instance provides a way of reducing the order of a linear differential equation at the expense of slightly more complex coefficients (in some sense transforming ”analytic” degrees of freedom into ”algebraic”). A difficult problem in the form of a high order linear partial differential equation then can be reduced to an ordinary differential equation or even an algebraic equation, which are typically considerably easier problems to solve. To this end, the capability to invert the Fourier transform is of course pivotal since this allows one to retrieve the solution of the difficult problem as an inverse Fourier transform of the solution to the simpler problem. The inverse Fourier transform can be formulated in terms of the Fourier transform itself, up to a constant and a reflection in the origin (of the domain), which is the content of the Fourier inversion theorem. This can be shown by applying the Fourier transform twice on a function, under some plausible assumptions that asserts that this is well-defined. However, it is not as straightforward as it may sound, and requires some effort to circumvent a lack of absolute convergence (i.e. integrability) in the iterated integrals.
that occur, which will be pointed out in the proof later. We now state the additional results which will be used. The way this is done here is essentially the same as the derivation in [5] which is based upon the concept of approximate identities and the following lemma. Two slightly different approaches, one also based on approximate identities and another which does not, of proving the Fourier inversion theorem can be found in [6].

**Lemma 2.12** (Approximate Identity Theorem). If \( f \in L^1(\mathbb{R}^n) \) is continuous at \( x \) and \( \{g_k(y)\} \) is a sequence of functions such that

(i) \( \int_{\mathbb{R}^n} g_k(y) \, dy = 1 \) \( \forall k \),

(ii) \( \lim_{k \to \infty} g_k(y) = 0 \) uniformly on \( y \in \mathbb{R}^n \setminus \mathcal{O} \) for every neighborhood \( \mathcal{O} \) of the origin, and

(iii) \( g_k(y) \geq 0 \) for all \( y \in \mathbb{R}^n \) and \( k \),

then

\[
\lim_{k \to \infty} g_k * f(x) = f(x)
\]

**Proof.** Let \( \epsilon > 0 \). Since \( f \) is continuous at \( x \), there is a neighborhood \( \mathcal{O}_x \) of \( x \) such that \( f(x) - \epsilon < f(y) < f(x) + \epsilon \) for all \( y \in \mathcal{O}_x \). From the third assumption we obtain

\[ g_k(x - y)(f(x) - \epsilon) < g_k(x - y)f(y) < g_k(x - y)(f(x) + \epsilon), \quad y \in \mathcal{O}_x \]

which implies, when integrating over \( \mathcal{O}_x \) w.r.t. \( y \),

\[
(f(x) - \epsilon) \int_{\mathcal{O}_x} g_k(x - y) \, dy \leq \int_{\mathcal{O}_x} g_k(x - y)f(y) \, dy \leq (f(x) + \epsilon) \int_{\mathcal{O}_x} g_k(x - y) \, dy
\]

(9)

Moreover, note that by the second assumption we have that for any neighborhood \( \mathcal{O} \) of 0,

\[
\lim_{k \to \infty} \int_{\mathcal{O}} g_k(y) \, dy = \lim_{k \to \infty} \left( \int_{\mathbb{R}^n} g_k(y) \, dy - \int_{\mathbb{R}^n \setminus \mathcal{O}} g_k(y) \, dy \right) = 1 - 0 = 1
\]

so that

\[
\lim_{k \to \infty} \int_{\mathcal{O}_x} g_k(x - y) \, dy = 1
\]

(10)
Secondly, let $\delta > 0$ be given and put $M = \int_{\mathbb{R}^n \setminus \mathcal{O}_x} |f(y)| \, dy$ if this integral is non-zero and $M = 1$ if it is 0. Note that the integral is always finite since $f \in L^1$. Then, choose $N$ large enough so that $g_k(x - y) \leq \frac{\delta}{M}$ on $\mathbb{R}^n \setminus \mathcal{O}_x$ for all $k > N$ (this is possible due to assumption (ii)). Then

$$\left| \int_{\mathbb{R}^n \setminus \mathcal{O}_x} g_k(x - y) f(y) \, dy \right| \leq \int_{\mathbb{R}^n \setminus \mathcal{O}_x} |g_k(x - y) f(y)| \, dy \leq \frac{\delta}{M} \int_{\mathbb{R}^n \setminus \mathcal{O}_x} |f(y)| \, dy < \delta$$

This shows that $\lim_{k \to \infty} \int_{\mathbb{R}^n \setminus \mathcal{O}_x} g_k(x - y) f(y) \, dy = 0$. Now, since

$$g_k * f(x) = \int_{\mathbb{R}^n \setminus \mathcal{O}_x} g_k(x - y) f(y) \, dy + \int_{\mathcal{O}_x} g_k(x - y) f(y) \, dy$$

we have

$$(f(x) - \epsilon) \int_{\mathcal{O}_x} g_k(x - y) \, dy \leq g_k * f(x) - \int_{\mathbb{R}^n \setminus \mathcal{O}_x} g_k(x - y) f(y) \, dy$$

$$\leq (f(x) + \epsilon) \int_{\mathcal{O}_x} g_k(x - y) \, dy$$

and by letting $k \to \infty$ we obtain

$$f(x) - \epsilon \leq \lim_{k \to \infty} g_k * f(x) \leq f(x) + \epsilon$$

The statement follows as $\epsilon > 0$ was arbitrary.

One may of course equivalently regard a parameter family $g_t(x)$ depending continuously on the parameter $t$, instead of a sequence. This is done in [6], and is also the way this fact is used in [5]. Such a family (or sequence) is typically called an approximate identity. The next result needed is about a particular approximate identity, which was encountered already in the first example.

**Proposition 2.13.** Let $\{a_k\}$ be a sequence of positive real numbers with $a_k \to 0$. Then the sequence $\{g_k(\xi)\} = \left\{ \frac{1}{(2\pi)^k} \mathcal{F}_x \left[ e^{-a_k |x|^2} \right] (\xi) \right\} = \left\{ \frac{1}{(2\pi)^n} \left( \sqrt{\frac{\pi}{a_k}} \right)^n e^{-|\xi|^2/4a_k} \right\}$ where $x, \xi \in \mathbb{R}^n$ is an approximate identity. In other words, it satisfies the criteria (i)-(iii) in Lemma 2.12.

*Proof.* First, positivity (iii) is obvious since the exponential function is strictly positive and the constants are positive. For (i), we have from $\int_{\mathbb{R}^n} e^{-ax^2} \, dx = \left( \sqrt{\frac{\pi}{a}} \right)^n$
that
\[
\int_{\mathbb{R}^n} e^{-\frac{\xi^2}{4a_k n}} d\xi = \left(\sqrt{4a_k \pi}\right)^n = \left(2\pi \sqrt{\frac{a_k}{\pi}}\right)^n = (2\pi)^n \left(\sqrt{\frac{\pi}{a_k}}\right)^{-n}
\]
which shows (i). For (ii), we note that since the exponential function is strictly increasing and \(\xi^2 = |\xi|^2 \geq 0\), we have that for any \(a > 0\), \(\sup_{|\xi| > r > 0} e^{-\frac{\xi^2}{4a n}} = e^{-\frac{r^2}{4a}}\). Now since \(\lim_{a \to 0^+} e^{-\frac{r^2}{4a}} \sqrt{\frac{\pi}{a}} = 0\) for any \(r > 0\) due to the exponential decay dominating over the growth of \(\sqrt{a^{-1}}\) as \(a \to 0^+\), we have that
\[
\lim_{a \to 0^+} \left(\sup_{|\xi| > r > 0} \left| e^{-\frac{\xi^2}{4a n}} \sqrt{\frac{\pi}{a}}\right|\right) = \lim_{a \to 0^+} \left| e^{-\frac{r^2}{4a}} \sqrt{\frac{\pi}{a}}\right| = 0
\]
It follows that \(g_k(\xi) \to 0\) as \(k \to \infty\) on any set \(\{\xi \in \mathbb{R}^n; |\xi| > r > 0\}\) and that the convergence is uniform. Since any neighborhood \(\mathcal{O}\) of 0 is \(\mathcal{O} = \{\xi \in \mathbb{R}^n; |\xi| < r\}\) for some \(r > 0\), any set \(\mathbb{R}^n \setminus \mathcal{O}\) is given by \(\{\xi \in \mathbb{R}^n; |\xi| \geq r > 0\}\) for some \(r > 0\), and we are done.

\[\square\]

As so often in analysis the Fourier inversion theorem is accessible by using a limiting argument. Due to the intricacies of interchanging integration and limits, especially when dealing with a more general case such as the class of \(L^1\) functions, this path to the Fourier inversion theorem requires a sufficiently strong tool to deal with the limits and integrals. It is here Lebesgue’s Theorem of Dominated Convergence (Theorem 2.6) is key, as it allows interchanging limits and integrals in the most general setting. With this, we are now ready to show the Fourier inversion formula.

**Theorem 2.14** (The action of \(F^2 = \mathcal{F}_\mathcal{F}\)). Suppose \(f\) is continuous and integrable, and that \(\mathcal{F} f = \hat{f}\) is integrable. Then
\[
\mathcal{F}(\mathcal{F} f(\xi))(x) = (2\pi)^n f(-x)
\]

**Proof.** We have
\[
\mathcal{F}(\mathcal{F} f(\xi))(x) = \mathcal{F}\left(\int_{\mathbb{R}^n} f(y) e^{iy\xi} dy\right) = \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} f(y) e^{iy\xi} dy\right) e^{i\xi x} d\xi
\]
As foretold, we immediately run into the issue that these iterated integrals are not absolutely convergent. Indeed, if one tries to trivially bound the expression using
the triangle inequality once or twice, one obtains either \(\left| \int_{\mathbb{R}^n} f(y) \, dy \right|\) or \(\int_{\mathbb{R}^n} |f(y)| \, dy\) as an integrand in the outer integral, which both are constants. Such an integral only converges if the constant is zero, i.e. if \(\left| \int_{\mathbb{R}^n} f(y) \, dy \right| = 0\) or \(\|f\|_{L^1} = 0\), but this of course is not the case in general. The latter would even imply \(f = 0\) almost everywhere. Owing to the lack of absolute convergence one can unfortunately not apply Fubini’s theorem to change the order of integration in this particularly important case. There is however a clever method to go around this. By multiplying the integrand by \(e^{-a\xi^2}\) for some \(a > 0\), it becomes absolutely convergent. We thus consider the following integral instead.

\[
\int_{\mathbb{R}^n} \left( \int_{\mathbb{R}^n} f(y) e^{-a\xi^2} e^{iy \cdot \xi} \, dy \right) e^{i\xi \cdot x} \, d\xi
\]

This is now a function of the introduced parameter \(a\) as well, in fact continuous, since \(e^{-a\xi^2}\) is continuous in \(a\). Now we may proceed to integrate in any order. First, integrating as it is,

\[
\int_{\mathbb{R}^n} \left( \int_{\mathbb{R}^n} f(y) e^{-a\xi^2} e^{iy \cdot \xi} \, dy \right) e^{i\xi \cdot x} \, d\xi = \int_{\mathbb{R}^n} \left( \int_{\mathbb{R}^n} f(y) e^{iy \cdot \xi} \, dy \right) e^{-a\xi^2} e^{i\xi \cdot x} \, d\xi
\]

Changing the order, we instead obtain

\[
\int_{\mathbb{R}^n} \int_{\mathbb{R}^n} f(y) e^{-a\xi^2} e^{iy \cdot \xi} \, dy \, e^{i\xi \cdot x} \, d\xi = \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} f(y) e^{-a\xi^2} e^{i(y+x) \cdot \xi} \, d\xi \, dy
\]

\[
= \int_{\mathbb{R}^n} f(y) \int_{\mathbb{R}^n} e^{-a\xi^2} e^{i(y+x) \cdot \xi} \, d\xi \, dy
\]

\[
= \int_{\mathbb{R}^n} f(z-x) \int_{\mathbb{R}^n} e^{-a\xi^2} e^{i(z-x) \cdot \xi} \, d\xi \, dz
\]

\[
(12)
\]

In (*) we have made that change of variables \(y \mapsto z = x + y\) in the outer integral. Note that, in the last expression in (12), we have introduced the notation \(If = I(f(t)) = f(-t)\) for composition with inversion through the origin of the domain, which henceforth will be referred to simply as inversion. The convolution itself is a
function of $x$.

For any $a > 0$, (11) and (12) must be equal by Fubini’s theorem. Now, we let $\{a_k\}$ be a positive sequence such that $a_k \to 0$ as $k \to \infty$. Then we must have, for each $k$,

$$
\int_{\mathbb{R}^n} \hat{f}(\xi) e^{-a_k \xi^2} e^{i\xi \cdot x} d\xi = (2\pi)^n (\mathcal{F}(2\pi)^{-n} e^{-a_k \xi^2}) (x)
$$

We now see that the integrand on the left hand side of (13), for every $k$, is dominated by $|\hat{f}(\xi) e^{i\xi \cdot x}| = |\hat{f}(\xi)|$ since $|e^{-a_k \xi^2}| \leq 1$, and this is integrable by assumption. Also, clearly $\hat{f}(\xi) e^{-a_k \xi^2} e^{i\xi \cdot x} \to \hat{f}(\xi) e^{i\xi \cdot x}$ as $k \to \infty$ since it is continuous in $a$. For the right hand side of (13), we see that the sequence $\{\mathcal{F}((2\pi)^{-n} e^{-a_k \xi^2})\}$ appearing in the convolution is an approximate identity from Proposition 2.13. Moreover, $f$ is continuous by assumption. Thus, by letting $k \to \infty$ in (13) we obtain by Lebesgue’s theorem of dominated convergence (Theorem 2.6) and by the approximate identity lemma (Lemma 2.12)

$$
\int_{\mathbb{R}^n} \hat{f}(\xi) e^{i\xi \cdot x} d\xi = (2\pi)^n f(-x)
$$

which is the desired result since the left hand side is $\mathcal{F} \hat{f} = \mathcal{F} \mathcal{F} f$. □

In other words, the Fourier transform applied twice on a function obeying a few reasonable assumptions is the inversion of the original function multiplied with $(2\pi)^n$. The inverse operation to the Fourier transform is therefore very similar to the Fourier transform itself. By noting that composition with inversion through the origin is an involution: $I(I(f)) = I(f(-x)) = f(-(-x)) = f(x)$, i.e. $I^{-1} = I$, we may simply state the equality in the above theorem as $\mathcal{F}(\mathcal{F}(f)) = (2\pi)^n I f$ which rearranges to $(2\pi)^{-n} I \mathcal{F}(\mathcal{F}(f)) = f$. Thus the inverse to $\mathcal{F}$ according to this is $(2\pi)^{-n} I \mathcal{F}$. Moreover, it is easy to check that $I \mathcal{F} f = \mathcal{F} I f$ simply by a change of sign of the integration variable $x \to -x$, and from this we also have that $\mathcal{F}((2\pi)^{-n} I \mathcal{F} f) = f$, i.e. the inverse works in both directions. This motivates the definition of the Inverse Fourier Transform $\mathcal{F}^{-1}$ as follows.

**Definition 2.15** (The Inverse Fourier Transform). We define the Inverse Fourier Transform of $f \in L^1$, denoted $\mathcal{F}^{-1} f$ or $\check{f}$, as

$$
\mathcal{F}^{-1} f(\xi) = \check{f}(\xi) = (2\pi)^{-n} I \mathcal{F} f = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} f(x) e^{-ix \cdot \xi} dx ; \xi \in \mathbb{R}^n
$$
For \( f \in L^2 \), we define it in the same way as above except with \( \mathcal{F} \) as the Fourier transform on \( L^2 \) as in Definition 2.3.

The coefficient is self-explanatory. The inversion composed with the Fourier transform amounts to the change of the sign in the exponent in the exponential factor, which is easily verified from with a change of variables. From the previous discussion, it follows from Theorem 2.14 that \( \mathcal{F} \) and \( \mathcal{F}^{-1} \) are inverse operations.

**Corollary 2.16** (The Fourier Inversion Theorem). With \( f \) and \( \widehat{f} \) as in Theorem 2.14,

\[
\mathcal{F}\mathcal{F}^{-1}f(x) = f(x) = \mathcal{F}^{-1}\mathcal{F}f(x)
\]

The Fourier Inversion theorem allows one to show a few other basic but important properties of \( \mathcal{F} \) and \( \mathcal{F}^{-1} \), in addition to those in proposition 2.10. \( \mathcal{F}^{-1} \) has in most cases the same properties as \( \mathcal{F} \) up to sign and a factor of \((2\pi)^{\pm n}\), since the inversion operation preserves most properties of \( \mathcal{F} \).

**Proposition 2.17** (Additional properties of \( \mathcal{F}, \mathcal{F}^{-1} \)). If \( f, g \in L^1 \) the following holds:

(i) \( \mathcal{F}^{-1}(c_1f(x) + c_2g(x)) = c_1\mathcal{F}^{-1}f(x) + c_2\mathcal{F}^{-1}g(x) \)

(ii) \( \mathcal{F}^{-1}f \) is bounded and continuous.

(iii) If \( f \in C^1 \), \( \frac{\partial}{\partial x_j}f \in L^1(\mathbb{R}^n) \) and \( f(x) \to 0 \) when \( |x_j| \to \infty \), then

\[
\mathcal{F}^{-1}\left( \frac{\partial f}{\partial x_j} \right)(\xi) = i\xi_j\mathcal{F}^{-1}f(\xi)
\]

(iv) If \( x_jf(x) \in L^1(\mathbb{R}^n) \) and \( x_j^2f(x) \in L^1(\mathbb{R}^n) \), then

\[
\mathcal{F}^{-1}(x_jf(x))(\xi) = i\xi_j\mathcal{F}^{-1}f(\xi)
\]

(v) \( \mathcal{F}^{-1}(f \ast g) = (2\pi)^n\mathcal{F}^{-1}f\mathcal{F}^{-1}g \)

Moreover, if \( f, g, \widehat{f}, \widehat{g}, \in L^1 \) and are all continuous, then

(vi) \( \mathcal{F}(fg) = (2\pi)^{-n}\mathcal{F}f \ast \mathcal{F}g \)

(vii) \( \mathcal{F}^{-1}(fg) = \mathcal{F}^{-1}f \ast \mathcal{F}^{-1}g \)
Proof. We omit most of the details for brevity as they for the most part are either trivial, have been shown previously or easily follows from previous results.

(i) is trivial (and practically identical to (i) in proposition 2.10).

(ii)-(v) follows from (ii)-(v) in proposition 2.10, Theorem 2.14 and the relations \( \mathcal{F} = (2\pi)^n I \mathcal{F}^{-1} \) by straightforward calculations.

We will show (vi) in detail, and then (vii) follows similarly. To this end, we apply \( \mathcal{F}^{-1} \) to the right hand side of (vi), and then use (i),(v) and Theorem 2.14 (Fourier inversion theorem, FIT) to obtain

\[
\mathcal{F}^{-1} \left( (2\pi)^{-n} \mathcal{F} f * \mathcal{F} g \right) \overset{(v)}{=} \mathcal{F}^{-1} (\mathcal{F} f) \mathcal{F}^{-1} (\mathcal{F} g) \overset{\text{FIT}}{=} fg
\]

Then (vi) follows by applying \( \mathcal{F} \) to both sides and applying Theorem 2.14 again.

Now we proceed by introducing a very important class of functions for our purposes, the Schwartz class, particularly appropriate in the setting of Fourier analysis and fundamental to its extension to distributions as will be seen. In all the results reviewed so far, as was pointed out in the very beginning, there are typically necessary conditions on the integrability, often not only on \( f \) but also on \( \hat{f} \). As noted in [6], one of the unfortunate facts in Fourier analysis is that there is in general no connection between the integrability of \( f \) and \( \hat{f} \), and typically a lot harsher conditions must be imposed on \( f \) to assert integrability of \( \hat{f} \). To make matters worse, if we want to also request differentiability of \( \hat{f} \), we see just from Proposition 2.10 that we must impose even harsher conditions not only on \( f \) but also on its partial derivatives, which goes beyond mere differentiability of \( f \). On the bright side, when inspecting this closely it is seen that the Fourier transform obeys a sort of conversion principle, which informally can be stated as converting smoothness of \( f \) into decay at infinity of \( \hat{f} \), and decay at infinity of \( f \) into smoothness of \( \hat{f} \), and vice versa. [5].

Formally, this should be interpreted as properties (ii), (iii) and (iv) in proposition 2.10 and their extension described in Remark 2.11, i.e. smoothness should be taken to be continuously differentiable with integrable partial derivatives (here we have
also assumed that they decay to 0 as $|x| \to \infty$ but this can be omitted); and decay at infinity should be interpreted as $x^n f \in L^1$, i.e. being able to control the growth of a monomial factor $x^n$ up to some degree $n$ to the extent that the integral of $x^n f(x)$ is finite. This can be stated in greater detail and generality, and there are also many more advanced and general results concerning this correspondence, most importantly the Riemann-Lebesgue Lemma and the many results referred to as Paley-Wiener theorems [5] - but we will again omit these for the sake of brevity and direct to the literature for details [5]. We note however that there is seemingly a loss in the conversion, in the sense that if $f$ has continuous and integrable partial derivatives up to order $k$, we only get decay of $\xi^n \hat{f}$ for monomial factors up to some degree $n < k$. Similarly, we proved integrability of $x^n f$ for $n$ up to 2 implied $\hat{f}$ is differentiable once. A question arising naturally from this observation is what the case would be under the assumption that $f \in C^\infty$ with all partial derivatives in $L^1$ and such that $x^n f \in L^1$ for every $n \geq 0$. In the definition of the Schwartz class, it is essentially in this way, except in a slightly more refined fashion, that this conversion principle is utilized and taken to its full length, which yields great results.

**Definition 2.18 (The Schwartz Class).** The Schwartz class $\mathcal{S}(\mathbb{R}^n)$ is the set of all functions $f : \mathbb{R}^n \to \mathbb{C}$ in $C^\infty$ with the property that for every multi-index $\alpha$,

$$
\lim_{|x| \to \infty} |x|^k |\frac{\partial^\alpha f}{\partial x^\alpha}(x)| = 0 \quad \forall k \in \mathbb{N} 
$$

(14)

If a function $f$ satisfies the limit criterion (14) for $|\alpha| = 0$ we call $f$ rapidly decreasing.

**Remark 2.19.**

(a) Note that the property of being rapidly decreasing can be rephrased to the equivalent statement that for every $n$, $\exists$ constants $r_n > 0, M_n \in \mathbb{R}$ such that $|f(x)| \leq M_n |x|^{-n}, |x| \geq r_n$, or equivalently $|x|^n |f(x)| \leq M_n, |x| \geq r_n$. Thus it is equivalent to saying that $|x|^n |f(x)|$ is (ultimately) bounded for every $n$. This is sometimes a more convenient form.
(b) The criterion \[ (14) \] for every \( \alpha \) is the same as requiring \( f \) and all of its partial derivatives of all orders to be rapidly decreasing.

In other words, the functions in \( S \) decreases more rapidly than any polynomial grows as \( x \to \infty \). It is nearly evident that the properties mentioned just before Definition \[ 2.18 \] is satisfied for \( f \in S \). In fact, something stronger holds as all these properties are also shared by all partial derivatives of \( f \). The basic properties of \( S \) are given below.

**Proposition 2.20** (Properties of \( S \)).

(i) \( S \) with pointwise multiplication is an algebra.

(ii) For any \( f \in S \) and any polynomial \( P \), \( P \left( \frac{\partial}{\partial x_1}, \cdots, \frac{\partial}{\partial x_n} \right) f \in S \).

(iii) For any \( f \in S \) and any polynomial \( P(x) \), \( P(x)f \in S \).

(iv) For any \( f \in S \) and \( k \geq 0 \) it holds that \( |x|^k f(x) \in L^p, 1 \leq p \leq \infty \). In particular, \( f \in L^p \) for all \( 1 \leq p \leq \infty \).

(v) For all \( f, g \in S \) we have \( f * g \in S \).

**Proof.**

(i) First, note that \( C^\infty \) is an algebra with pointwise multiplication. It remains to show that \[ (14) \] is preserved under the operations of an algebra. Closure with respect to scalar multiplication and addition follows from the basic properties of limits and the triangle inequality since this implies that the limit criterion \[ (14) \] is preserved under such operations. By the product rule, for any multi-index \( \alpha \),

\[
\frac{\partial^\alpha (fg)}{\partial x^\alpha} = \sum_{i=1}^{N} c_i \frac{\partial^{\beta_i} f}{\partial x^{\beta_i}} \cdot \frac{\partial^{\gamma_i} g}{\partial x^{\gamma_i}}
\]

for some constants \( c_i \) and multi-indices \( \beta_i, \gamma_i \). For every \( k \geq 0 \), by the triangle inequality

\[
|x|^k \left| \frac{\partial^\alpha (fg)}{\partial x^\alpha} \right| \leq \sum_{i=1}^{N} |c_i| |x|^k \left| \frac{\partial^{\beta_i} f}{\partial x^{\beta_i}} \right| \left| \frac{\partial^{\gamma_i} g}{\partial x^{\gamma_i}} \right| \to 0
\]
as $|x| \to \infty$, since for each $i$, $|x|^k \left| \frac{\partial^k f}{\partial x^k} \right| \to 0$ and $\left| \frac{\partial^i g}{\partial x^i} \right| \to 0$ as $|x| \to \infty$ from $f, g \in S$ and (14).

(ii) It suffices to show that $f \in S$ implies $\frac{\partial f}{\partial x_j} \in S$; the statement then follows by induction and the vector space property of $S$ from (i). It is clear that since $f \in C^\infty$, then $\frac{\partial f}{\partial x_j} \in C^\infty$. Trivially, every derivative of $\frac{\partial f}{\partial x_j}$ is also a derivative of $f$, so that it must be rapidly decreasing. Hence $\frac{\partial f}{\partial x_j} \in S$ for any $x_j$.

(iii) Since all polynomials are $C^\infty$ and this is an algebra, clearly $P(x)f(x)$ is also $C^\infty$. For any polynomial $P(x)$, $|x|^k P(x)f(x)$ is a sum consisting of terms of the form $C_k |x|^k x_1^{p_1} \ldots x_n^{p_n} f(x)$. We have that $|x| \geq |x_j| \geq x_j$ for all $j$, and it follows that $C_k |x|^k x_1^{p_1} \ldots x_n^{p_n} f(x) \leq C_k |x|^k |x|^{p_1} \ldots |x|^{p_n} f(x) = C_k |x|^{k+p_1+\ldots+p_n} f(x)$ which tends to zero as $|x| \to \infty$ for any positive integer $k$. The sum of these terms must also tend to zero in this limit for any $k$, so $P(x)f(x)$ is rapidly decreasing. To see that any partial derivative of $P(x)f(x)$ is rapidly decreasing, we simply note that any partial derivative of $P(x)$ is a polynomial, and $S$ is closed with respect to differentiation by (ii), hence every term in the sum

$$\frac{\partial^\alpha P(x)f(x)}{\partial x^\alpha} = \sum_{i=1}^{N} P_i(x) \frac{\partial^{\beta_i} f}{\partial x^{\beta_i}}$$

for an arbitrary multi-index $\alpha$ is a product of the form $P_i(x)g_i(x)$ with $P_i$ a polynomial and $g_i \in S$, and therefore is rapidly decreasing by what we have just shown.

(iv) We omit the details but mention that this can be showed by utilizing that $|x|^k |f(x)|$ for $f \in S$ is ultimately bounded by $|x|^{-(n+1+k)} \in L^1(\mathbb{R}^n \setminus \mathcal{O})$ (where $\mathcal{O}$ is some fixed neighbourhood of the origin). By dividing the integral into two parts, one over a set $\{|x| > r > 0\}$ where this bound is valid and another over the remaining compact hyper-disk $\{|x| \leq r\}$ on which $|f(x)|$ is
bounded, one can show that both are finite. For the general $L^p$-case with $1 \leq p < \infty$, the method is precisely the same except that the bound is chosen appropriately differently, i.e. the power of the polynomial denominator is chosen with respect to $p$ to achieve a finite $L^p$-norm. For $p = \infty$, we have that $f$ is bounded on compact sets due to continuity, and becomes uniformly small outside some compact set due to (14). Hence $f \in L^\infty$ as it is bounded.

(v) $f \ast g$ is $C^\infty$ due to the properties of the convolution (Proposition 2.9 (iii)) and the facts that $f, g \in C^\infty$ and all of their partial derivatives are integrable (and even square-integrable) by (iv).

To show (14), we have

$$|x|^k |f \ast g(x)| = |x|^k \left| \int_{\mathbb{R}^n} f(x - y)g(y) \, dy \right|$$

$$\leq \int_{\mathbb{R}^n} (|x - y| + |y|)^k |f(x - y)g(y)| \, dy$$

$$= \sum_{i=0}^{k} \binom{k}{i} \int_{\mathbb{R}^n} |x - y|^i |f(x - y)||y|^{k-i} |g(y)| \, dy$$

by the triangle inequality and binomial theorem. Thus, if every integral in the sum is finite the statement follows, by Remark 2.19. For a general such integral with arbitrary non-negative integers $p, q$, we have

$$\int_{\mathbb{R}^n} |x - y|^p |f(x - y)||y|^q |g(y)| \, dy \leq \int_{\mathbb{R}^n} \sup_{y \in \mathbb{R}^n} (|x - y|^p |f(x - y)||y|^q |g(y)|) \, dy$$

$$= \sup_{x \in \mathbb{R}^n} (|x|^p |f(x)||y|^q |g(y)|) \int_{\mathbb{R}^n} |y|^q |g(y)| \, dy$$

$$= \| |x|^p f \|_{L^\infty} \| |y|^q g \|_{L^1} < \infty$$

since $|x - y|^p |f(x - y)|$ is bounded and $|y|^q |g(y)|$ is integrable from (iv) above, for any $p$ and $q$.

Finally, to see that all partial derivatives of $f \ast g$ are rapidly decreasing, we simply note that for any multi-index $\alpha$,

$$\frac{\partial^\alpha f \ast g}{\partial x^\alpha}(x) = \frac{\partial^\alpha f}{\partial x^\alpha} \ast g(x)$$
and $S$ is closed with respect to differentiation by (ii), so $\frac{\partial f}{\partial x} * g$ is again a convolution between two functions in $S$ and thus must be rapidly decreasing by the previous argument.

$\blacksquare$

The reason why the Schwartz class is particularly interesting is however not essentially the above properties, or even that such functions have enough properties for all mentioned results to hold. In fact, for many of these results, the Schwartz class have a lot more than necessary. The reason is instead mainly the following fact.

**Theorem 2.21** (Closure of $S$ under $F$ and $F^{-1}$). For $f \in S$, it holds that $F f \in S$ and $F^{-1} f \in S$. In other words, $F : S \to S$ is a bijection with inverse $F^{-1}$.

**Proof.** We need to show that $f \in S$ implies that $F f \in C^\infty$ and that all its partial derivatives are rapidly decreasing. First, to see that $\hat{f} \in C^\infty$, we have for any polynomial $p$,

$$p \left( \frac{\partial}{\partial \xi_1}, \ldots, \frac{\partial}{\partial \xi_n} \right) \hat{f}(\xi) = F \left( p(-ix)f(x) \right)(\xi)$$

where $p(-ix)f(x) \in S \subset L^1$, so that the right-hand side is continuous from Proposition 2.10 (ii). Thus every partial derivative, of any order, of $\hat{f}$ is continuous.

Next, to show rapid decrease, let $p, q$ be polynomials and let $p^\dagger, q^\dagger$ be the polynomials such that $p(x) = p^\dagger(-ix)$ and $q(x) = q^\dagger(-ix)$ (such $p^\dagger, q^\dagger$ always exist, and is obtained by multiplying each coefficient of $p$ and $q$ respectively with $1, i, -1$ or $-i$ depending on the order of the term mod 4). Then, by Proposition 2.10

$$p(\xi) q \left( \frac{\partial}{\partial \xi} \right) \hat{f}(\xi) = p^\dagger(-i\xi) q^\dagger \left( -i \frac{\partial}{\partial \xi} \right) \hat{f}(\xi)$$

$$= p^\dagger(-i\xi) F \left( q^\dagger f \right)(\xi)$$

$$= F \left( p^\dagger \left( \frac{\partial}{\partial x} \right) (q^\dagger f) \right)(\xi)$$

$$= \hat{f^\dagger}(\xi)$$

where $f^\dagger = p^\dagger \left( \frac{\partial}{\partial x} \right) (q^\dagger f)$, which holds since $f, q^\dagger f, f^\dagger \in S$ by assumption and from Proposition 2.20. Moreover, $\hat{f^\dagger}$ is continuous and bounded, hence so is $p(\xi) q \left( \frac{\partial}{\partial \xi} \right) \hat{f}(\xi)$.
for any \( p, q \). For \( k \in \mathbb{N} \), put \( p(\xi) = (\xi_1^2 + \cdots + \xi_n^2)^k \). Then \( |\xi|^k \leq p(\xi) \) for \( |\xi| \geq 1 \), so that

\[
|\xi|^k \left| q \left( \frac{\partial}{\partial \xi} \right) \hat{f}(\xi) \right| \leq p(\xi) q \left( \frac{\partial}{\partial \xi} \right) \hat{f}(\xi) \leq M_{q,k}, \quad |\xi| \geq 1
\]

for some \( M_{q,k} \in \mathbb{R} \). Since \( q \) and \( k \) was arbitrary, it follows from Remark 2.19 that \( \frac{\partial |\alpha|}{\partial \xi^\alpha} \hat{f} \) is rapidly decreasing for any multi-index \( \alpha \).

Lastly, it suffices to to show that inversion, i.e. \( f(x) \rightarrow f(-x) \) preserves these properties; it then follows that also \( \mathcal{F}^{-1} \) satisfies the claims, as it is composed of inversion, \( \mathcal{F} \) and multiplication with a constant. The property of being rapidly decreasing clearly is preserved since \( |x| = |x| \), so that if \( |f(x)| \leq M_k |x|^{-k} \) for \( |x| > R \), then evidently \( |f(-x)| \leq M_k |x|^{-k} \) for \( |x| > R \). Moreover, \( \frac{\partial}{\partial x_j}(f(-x)) = -\frac{\partial f}{\partial x_j}(-x) \) and inversion preserves continuity. It follows that it preserves the property of being \( C^\infty \) and rapid decrease of all partial derivatives. Thus \( f(-x) \in \mathcal{S} \) whenever \( f(x) \in \mathcal{S} \), and so \( \mathcal{F}^{-1} f = (2\pi)^n I(\mathcal{F} f) \in \mathcal{S} \) for all \( f \in \mathcal{S} \). \( \square \)

To conclude this section, we aim to state and prove the Plancherel’s formula for functions in the Schwartz Class. Plancherel’s formula is a classical and central result in Fourier analysis and it’s application to partial differential equations, and it will see extensive use in the last part of this thesis. In many applications of partial differential equations, the \( L^2 \)-norm of the solutions represents important quantities, notable examples being energy and probability in physics. Plancherel’s formula can be used to prove conservation or bounds for these quantities, without requiring explicit solutions.

First we prove a result for \( L^1 \), and hence for \( \mathcal{S} \), which will facilitate the proof and which will also be necessary when defining the Fourier transform of distributions.

**Proposition 2.22.** If \( f, g \in L^1(\mathbb{R}^n) \), then

\[
\int_{\mathbb{R}^n} \hat{f} \hat{g} = \int_{\mathbb{R}^n} f \hat{g}
\]
Proof. The proof is simple if we can motivate a change in order of integration. In that case, we have
\[ \int_{\mathbb{R}^n} \hat{f}(\xi)g(\xi) \, d\xi = \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} g(\xi)f(x)e^{ix\cdot\xi} \, dx \, d\xi \]
\[ \overset{(*)}{=} \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} f(x)g(\xi)e^{i\xi\cdot x} \, d\xi \, dx \]
\[ = \int_{\mathbb{R}^n} f(x)\hat{g}(x) \, dx \]
which would be the complete proof, if we can justify \((*)\). To this end, we have that the iterated integral satisfies
\[ \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} |g(\xi)||f(x)| \, dx \, d\xi = \left( \int_{\mathbb{R}^n} |f(x)| \, dx \right) \left( \int_{\mathbb{R}^n} |g(\xi)| \, d\xi \right) = \|f\|_{L^1} \|g\|_{L^1} < \infty \]
By Tonelli’s theorem, this is enough to ensure that \(|g(\xi)f(x)|\) is integrable on \(\mathbb{R}^n \times \mathbb{R}^n\), and therefore \(g(\xi)f(x)e^{ix\cdot\xi} \in L^1(\mathbb{R}^n \times \mathbb{R}^n)\), and Fubini’s theorem then confirms that the different iterations of the integrals are equal. This completes the proof. □

This of course applies to \(S\) since we have shown that \(S \subseteq L^1\). Note that Plancherel’s formula is valid for \(L^2\), and the previous result can also be extended to \(L^2\), relying on the fact that \(L^1 \cap L^2\) is \(L^2\)-dense in \(L^2\). In fact, we have the following, which we will not prove here, but will rely on in the following proof.

**Theorem 2.23.** \(S\) is dense in \(L^2\), in \(L^2\)-sense.

Now we finish by proving Plancherel’s formula.

**Theorem 2.24 (Plancherel’s formula).** If \(f \in L^2(\mathbb{R}^n)\) then \(\hat{f} \in L^2(\mathbb{R}^n)\) and
\[ (2\pi)^n \int_{\mathbb{R}^n} |f(x)|^2 \, dx = \int_{\mathbb{R}^n} \left| \hat{f}(\xi) \right|^2 \, d\xi \]

**Proof.** By the definition of the Fourier transform of an \(L^2\)-function, it is enough to show the statement when \(f \in L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)\), and in view of Theorem 2.23 it suffices to show for \(S(\mathbb{R}^n)\). The statement then extends to \(L^2\) by a limit argument. First, note that
\[ (\mathcal{F}(f))^* = \left( \int_{\mathbb{R}^n} f(x)e^{ix\cdot\xi} \, dx \right)^* = \int_{\mathbb{R}^n} f^*(x)e^{-ix\cdot\xi} \, dx = (2\pi)^n \mathcal{F}^{-1}(f^*) \]
which follows from the definitions of $F$ and $F^{-1}$.

Now, since $|f(x)|^2 = f(x)f^*(x)$ for any $f$, with the star denoting complex conjugates, we have
\[
\int_{\mathbb{R}^n} |Ff(x)|^2 \, dx = \int_{\mathbb{R}^n} Ff(x)(Ff(x))^* \, dx \overset{\text{Eq. (15)}}{=} (2\pi)^n \int_{\mathbb{R}^n} Ff(x)F^{-1}(f^*)(x) \, dx = (2\pi)^n \int_{\mathbb{R}^n} f(x)(f(x))^* \, dx = (2\pi)^n \int_{\mathbb{R}^n} |f(x)|^2 \, dx
\]
which proves the theorem for $f \in S$.

To show the statement for general $f \in L^2$, we may in view of Theorem 2.23 pick a sequence $\phi_k$ in $S$ such that $\|\phi_k - f\|_{L^2} \to 0$ as $k \to \infty$. Since $\{\phi_k\} \subset L^1 \cap L^2$, by the definition of the Fourier transform on $L^2$, we have that
\[
\hat{f}(\xi) = \lim_{k \to \infty} \hat{\phi}_k(\xi)
\]
where the convergence is in $L^2$. In fact, we have that for each $k, l$, $\hat{\phi}_k - \hat{\phi}_l \in S$, and hence by Plancherel’s formula,
\[
\|\hat{\phi}_k - \hat{\phi}_l\|_{L^2} = \|(2\pi)^n F^{-1}(\hat{\phi}_k - \hat{\phi}_l)\|_{L^2} = (2\pi)^\frac{n}{2} \|\phi_k - \phi_l\|_{L^2}
\]
which shows that since $\{\phi_k\}$ is Cauchy, so must $\{\hat{\phi}_k\}$ be, and then by completeness of $L^2$ in the $L^2$-norm, $\hat{\phi}_k$ converges to some element $\hat{f} \in L^2$. Lastly we then have by Plancherel’s formula for $S$,
\[
\|\hat{f}\|_{L^2} = \lim_{k \to \infty} \|\hat{\phi}_k\|_{L^2} = \lim_{k \to \infty} \|(2\pi)^n \phi_k\|_{L^2} = \|(2\pi)^n f\|_{L^2}
\]
This completes the proof of the theorem for $L^2$.

\[\Box\]

We conclude this initial section with some comments on Plancherel’s formula. In some sense, Plancherel’s formula completes the definition of the Fourier transform on $L^2$, because it shows that every $L^2$ function indeed has a Fourier transform that is a function also in $L^2$. In fact it shows that $F: L^2 \to L^2$ is a bijection and $(2\pi)^{-\frac{n}{2}} F$ is an isometry (i.e. norm preserving linear map) on $L^2$. This is remarkable in the sense that at first glance, the integral defining $F$ only obviously exists definitely on $L^1$, and outside of $L^1$ things are rather unclear. But as it turns out, it also is well-defined
on the entirety of $L^2$. $L^2$ is a rather large, general space, containing functions that can behave wildly, and boundedness and continuity is far from granted. In contrast, we saw $S \subset L^2$ as an example of a much smaller subspace of smooth and bounded functions that is also preserved by $\mathcal{F}$. As hinted, the true importance of $S$ will show in the next section.

3. Distributions

Now we turn our attention to distributions, with the ultimate goal of extending differentiation and the Fourier transform to these. The central idea in distribution theory is that instead of studying real- or complex valued functions, one considers certain functionals - that is, linear maps from a vector space into the underlying field on function spaces of real- and complex valued functions over the real- or complex numbers. In terms of linear algebra and functional analysis, these are the elements of the dual space. The underlying space of functions is of outmost fundamental importance, because the general approach when extending analytical concepts and properties to distributions is by doing so via the functions in this space - many properties of the functions can in this manner be defined and transferred to the functionals. The functions in the underlying vector space are in distribution theory called test functions, and the distributions are the functionals in the dual space. The particular space chosen as the test function space, together with a topology, determines the dual space, and in this way defines the type of distributions. Here, we first introduce the general class of distributions and later we will consider the subclass of tempered distributions. To begin with, we must define the space of test functions and for this we remind of the notion of support of a function.

**Definition 3.1** (Support of a function). The support of a real or complex valued function $f : \Omega \to \mathbb{C}$, $\Omega \subset \mathbb{R}^n$ is the set

$$\text{supp } f := \{x \in \Omega : f(x) \neq 0\}$$

where the bar denotes the closure of the set, i.e. the smallest closed set containing the set in question.
Now we introduce the initial space of test functions for general distributions: the space $C_c^\infty$ of smooth functions with compact support.

**Definition 3.2 (The space $C_c^\infty$, test functions).** Let $\Omega \subset \mathbb{R}^n$ be open. $C_c^\infty(\Omega)$ is then the class consisting of all complex-valued functions $\varphi$ on $\Omega$ fulfilling

(a) $\varphi \in C^\infty(\Omega)$,

(b) $\text{supp } \varphi$ is a compact subset of $\Omega$.

The second criterion (b) can alternatively be stated as requiring $\text{supp } \varphi$ to be bounded in the case $\Omega = \mathbb{R}^n$. The reason why we impose (b) will be apparent first when we define distributions properly, and when we establish the identification of functions as distributions. For future reference, we note that the important implication is that $\varphi \in C_c^\infty$ will be zero in some neighborhood of any point of $\partial \Omega$ thus enabling to "kill" any other reasonable function $f$, in the sense that $f(x)\varphi(x) \to 0$ when approaching the boundary. Moreover, any integral of $f(x)\varphi(x)$ will reduce to be over a compact subset of $\Omega$ instead. Hence the behavior of $f$ close to the boundary $\partial \Omega$ is in principle irrelevant for products $f(x)\varphi(x)$ and especially when integrating these.

Secondly, note that $C_c^\infty$ is a vector space, since $C^\infty$ is a vector space, and the criterion (b) is also preserved under linear combinations. To see this, note that multiplication by a scalar does not alter $\text{supp } \varphi$. When taking sums, we have that $\text{supp}(\varphi_1 + \varphi_2) \subset (\text{supp } \varphi_1 \cup \text{supp } \varphi_2)$ and a finite union of compact subsets of $\Omega$ is of course also a compact subset of $\Omega$. Building on to this, we may consider a product $\varphi_1 \varphi_2$ as well. This is non-zero if and only if $\varphi_1(x)$, $\varphi_2(x)$ both are non-zero, which requires $x \in \text{supp}(\varphi_1) \cap \text{supp}(\varphi_2)$ and so we have $\text{supp}(\varphi_1 \varphi_2) = \text{supp}(\varphi_1) \cap \text{supp}(\varphi_2)$. This is a compact subset of $\Omega$, so that $\varphi_1 \varphi_2 \in C_c^\infty$ since $\varphi_1, \varphi_2 \in C^\infty$ as well. $C_c^\infty$ is hence an algebra. This implies something even stronger however, namely that $C_c^\infty$ is closed with respect to multiplication with any $C^\infty$ function, since $\text{supp}(\varphi_1 \varphi_2)$ is compact if at least one of $\varphi_1$, $\varphi_2$ has compact support.
Finally one may also note that this class is quite restrictive. In fact it is easy to see that no polynomial (other than the zero polynomial) is in $C_c^\infty$, nor are $e^x$, $\sin x$ or $\cos x$. $C_c^\infty$ is more restrictive than $S$, which can be seen by noting that $e^{-x^2} \in S$ but this function is never zero so that $e^{-x^2} \notin C_c^\infty$. On the other hand, we have $C_c^\infty \subset S$ since all partial derivatives of a function in $C_c^\infty$ and the function itself is zero outside a bounded set and thus must be rapidly decreasing. One difficult part when constructing a function in $C_c^\infty$ is that one must ensure that, for some compact subset $A \subset \Omega$, the derivatives of $f$ of all orders must tend to zero when approaching $\partial A$. From this we can also realize that $C_c^\infty$ is closed with respect to differentiation, since if $\varphi$ is smooth and vanishes outside a bounded set $B$ then $\frac{\partial^n \varphi}{\partial x^\alpha}$ is also smooth and must vanish outside of $B$.

We summarize this discussion in a proposition.

**Proposition 3.3** (Properties of $C_c^\infty$). Let $\Omega \subset \mathbb{R}^n$ be open. Then

(i) $C_c^\infty(\Omega)$ is an algebra with pointwise multiplication,

(ii) $C_c^\infty(\mathbb{R})^n \subset S(\mathbb{R}^n)$,

(iii) For any $\varphi \in C_c^\infty(\Omega)$ and $f \in C^\infty(\Omega)$, it holds that $f \varphi \in C_c^\infty(\Omega)$,

(iv) For any $\varphi \in C_c^\infty(\Omega)$, $P\left(\frac{\partial}{\partial x_1}, \cdots, \frac{\partial}{\partial x_n}\right) \varphi(x) \in C_c^\infty(\Omega)$ for any polynomial $P$.

One may note that the second statement of course induces all properties from $S$ from Proposition 2.20 in the previous section, in particular integrability.

To define distributions, a notion of continuity is needed and in order to obtain this one defines a mode of convergence in $C_c^\infty$. To facilitate this we first remind of the sup-norm.

**Definition 3.4** (The sup-norm, $\|\cdot\|_\infty$). For a function $f : \Omega \to \mathbb{C}$ on a set $\Omega \subset \mathbb{R}^n$, the supremum norm or sup-norm of $f$ on $\Omega$, denoted $\|f\|_{\infty, \Omega}$ is

$$\|f\|_{\infty, \Omega} = \sup_{x \in \Omega} |f(x)|$$

We write $\|\cdot\|_{\infty, \Omega} = \|\cdot\|_\infty$, whenever $\Omega$ is irrelevant or obvious from the context.
This is a norm in the ordinary sense on any space of bounded functions, i.e. it satisfies positivity, that only the zero function has $\|f\|_\infty = 0$, and the triangle inequality. $\|-\|_\infty$ is thus defined on $C_c^\infty$ since all functions in this set are bounded. Also, convergence in the sup-norm on $\Omega$ implies uniform convergence on the set $\Omega$.

Now we may define convergence in $C_c^\infty$.

**Definition 3.5** (Convergence of test functions in $C_c^\infty(\Omega)$). Let $\{\varphi_n\}$ be a sequence in $C_c^\infty(\Omega)$. We say that $\varphi_n$ converges to $\varphi$ in $C_c^\infty(\Omega)$ or as test functions, written $\varphi_n \to \varphi$ or $\lim_{k \to \infty} \varphi_n = \varphi$ if

1. for every multi-index $\alpha$ it holds that
   $$\lim_{k \to \infty} \left\| \frac{\partial^\alpha}{\partial x^\alpha} \varphi_n - \frac{\partial^\alpha}{\partial x^\alpha} \varphi \right\|_{\infty, \Omega} = 0$$
2. There exist an $R > 0$ such that $\text{supp} (\varphi) \subset \{x \in \mathbb{R}^n; |x| < R\}$ and $\text{supp}(\varphi_n) \subset \{x \in \mathbb{R}^n; |x| < R\}$ for all $n$. That is, the supports of the functions $\varphi, \varphi_n$ is uniformly bounded with respect to $n$.

**Remark 3.6.** This kind of convergence is very strong. First of all, it implies that any sequence of partial derivatives of $\varphi_n$ (with the same multi-index $\alpha$) converges uniformly on $\Omega$, which in turn implies that continuity is conserved through the limit, so that the limit function must also be $C^\infty$. The second criterion asserts that the support of the limit function also must be bounded. If $\Omega = \mathbb{R}^n$, this means that the limit function must lie in $C_c^\infty(\mathbb{R}^n)$, i.e. $C_c^\infty(\mathbb{R}^n)$ is closed with respect to this mode of convergence.

Distributions on $\Omega$ are now taken to be the elements of the dual of $C_c^\infty(\Omega)$ with this topology, more precisely as follows.

**Definition 3.7** (Distributions, $\mathcal{D}'(\Omega)$). A distribution on $\Omega$ is a linear and continuous map $\langle f, - \rangle : C_c^\infty(\Omega) \to \mathbb{C}$, $\varphi \mapsto \langle f, \varphi \rangle$. That is, $\langle f, - \rangle$ satisfies
(1) (Linearity) \( \langle f, c_1 \varphi_1 + c_2 \varphi_2 \rangle = c_1 \langle f, \varphi_1 \rangle + c_2 \langle f, \varphi_2 \rangle \)
for all \( \varphi_1, \varphi_2 \in C^\infty_c(\Omega) \) and \( c_1, c_2 \in \mathbb{R} \).

(2) (Continuity) \( \lim_{k \to \infty} \langle f, \varphi_n \rangle = \langle f, \varphi \rangle \) whenever \( \varphi_n \to \varphi \) in \( C^\infty_c(\Omega) \)-sense.

The set of all distributions on \( \Omega \) is denoted \( D'(\Omega) \).

Distributions are considered generalized functions, indicating that ordinary functions also should be distributions. We should note that it is the notion of real- or complex-valued function in mathematical analysis that is referred to here, and of course not the set theoretic notion of function. At first glance it seems as if there are many ways one could associate a function to a distribution. One simple way would be evaluation, \( \langle f, \psi \rangle = f(a)\psi(a) \) for some \( a \). This is clearly linear, and also continuous. However, it is simply not interesting at all since it only involves evaluation at a single point and ignores all other properties of the function as a whole, amounting to an almost complete loss of information. It involves an arbitrary choice of \( a \), and thus there are many possible distributions which would correspond to a single function \( f \) in this way. Moreover, if two functions coincide only in the single point \( a \), they will in this manner define the same distribution, even if they are wildly different otherwise. In fact, such a distribution may as well be written \( \langle T, \psi \rangle = C\psi(a) \) for some constant \( C \). One can show that every evaluation distribution of this type may be expressed using a single distribution, the famous Dirac delta distribution \( \delta \), which will be introduced in a moment. There are indeed much more interesting ways a function may be viewed as a distribution which takes into account much more of the behaviour and properties of the function, and not surprisingly, the standard way is such. It is based on integration. To see how this can be done, observe that integration over a fixed set is in itself a linear functional of functions. The next observation is then that \( \int_{\Omega} f(x)\varphi(x) \, dx \) is also a linear functional of \( \varphi \) for a fixed function \( f \), as long as the integral exists for all considered functions \( \varphi \), that is for \( \varphi \in C^\infty_c(\Omega) \) in our case. For this to be true, we need the following condition on \( f \).
**Definition 3.8** (Local Integrability). We say that a function \( f : \Omega \to \mathbb{C}, \Omega \subset \mathbb{R}^n \) is locally integrable, and write \( f \in L^1_{\text{loc}}(\Omega) \), if

\[
\int_M |f(x)| \, dx < \infty
\]

for every compact subset \( M \subset \Omega \).

Local integrability is a sufficient condition for the integral \( \int_{\Omega} f(x)\varphi(x) \, dx \) to be finite for every \( \varphi \in C^\infty_c(\Omega) \), because every such \( \varphi \) is identically zero outside the compact subset of \( \Omega \) which is \( \text{supp}(\varphi) \). The integral thus reduces to be over such a compact subset, on which \( \varphi \) must be bounded, so that \( \int_{\Omega} |f(x)\varphi(x)| \, dx \leq M \int_{\text{supp}(\varphi)} |f(x)| \, dx < \infty \) given that \( f \in L^1_{\text{loc}}(\Omega) \). In other words, for \( f \in L^1_{\text{loc}}(\Omega) \),

\[
\langle f, \varphi \rangle = \int_{\Omega} f(x)\varphi(x) \, dx
\]

is a functional defined on the whole of \( C^\infty_c(\Omega) \). It is additionally easily seen to be linear, and can be shown to be continuous in the distributional sense (\( \mathcal{D}' \)-sense), so that this is a distribution in \( \mathcal{D}'(\Omega) \) according to the above definition. One thus identify locally integrable functions as distribution in the following way.

**Definition 3.9** (Functions in \( L^1_{\text{loc}} \) as a distributions). For \( f \in L^1_{\text{loc}}(\Omega) \), we define the distribution corresponding to \( f \), denoted \( \langle f, - \rangle \), \( T_f \) or \( \langle T_f, - \rangle \), as

\[
\langle f, \varphi \rangle := \int_{\Omega} f(x)\varphi(x) \, dx
\]

for \( \varphi \in C^\infty_c(\Omega) \).

**Proposition 3.10.** With \( f \) and \( \langle f, - \rangle \) as in the above definition, \( \langle f, - \rangle \in \mathcal{D}'(\Omega) \).

**Proof.** We have already shown that it is defined for every \( \varphi \in C^\infty_c(\Omega) \). It remains to show the required properties. Linearity is trivial, and follows from linearity of the integral. Only continuity remains.
(2) Continuity: Let $\varphi_k \to \varphi$ in $C^\infty_c(\Omega)$-sense. Then there is a compact set $B \subset \Omega$ containing the supports of $\varphi$ and all $\varphi_k$, and thus

$$
\lim_{k \to \infty} \langle f, \varphi_k \rangle = \lim_{k \to \infty} \int_{\Omega} f(x)\varphi_k(x) \, dx
= \lim_{k \to \infty} \int_{B} f(x)\varphi_k(x) \, dx
= \int_{B} \left( \lim_{k \to \infty} f(x)\varphi_k(x) \right) \, dx
= \int_{\Omega} f(x)\varphi(x) \, dx = \langle f, \varphi \rangle
$$

We may interchange the integral and the limit because the convergence is uniform on $B$ and $f$ is integrable over $B$.

\[ \square \]

Still, we may ask how exactly this generalizes a real- or complex valued function on $\mathbb{R}^n$; distributions clearly doesn’t take inputs from $\mathbb{R}^n$. The main point is that precise pointwise information about functions in surprisingly many situations is superfluous and irrelevant. In $L^p$-spaces, one does not consider functions who differ on a set of measure zero, let alone at individual points, as different. Closely related is also the question whether one should insist that a solution to a differential equation really have to be differentiable at exactly every point. In many applications there is no reason for this. Passing from functions to distributions in distribution theory in some sense involves giving up the unimportant pointwise information of functions on negligibly small sets, i.e. with measure zero, to focus solely on the behaviour of the functions at large.

The identification of functions as distributions in Definition 3.9 can be interpreted as an averaging process. Instead of pointwise values $f(x)$, $f$ is viewed in terms of its averages $\int_{\Omega} f(x)\varphi(x) \, dx$ where the test functions $\varphi$ take the roles of weights in the averages. In this way, the distinct pointwise information of $f$ is smoothed out and replaced by the local average behaviour which captures the essence of $f$, void of negligible details. If we where to change the values of $f$ in only a few points, it would strictly speaking yield a different function $g$, yet which mostly would have the same properties as $f$: when integrating they would yield identical results, and if
$f$ satisfies a differential equation, so would $g$ except possibly at the points where it differs from $f$. The argument is then that not much is to be gained from considering $f$ and $g$ to be different. However, the averages $\int_{\Omega} f(x)\varphi(x) \, dx$ would not change from this, i.e. $f = g$ as distributions.

The averaging analogy can in an illustrative manner be taken even further by comparing to the measurement of physical quantities, as is done by Strichartz [5]. When measuring a physical variable/quantity $f$, we can represent the measuring process with a weight function $\varphi$ which is the amalgam of the conditions of the system we measure at the time of the measurement, the accuracy of the method used etc. The outcome of the measurement is not the value $f(x)$, but rather an average, sometimes in several different aspects simultaneously. It could for instance be a spatial average of the system, the average over a period of time or the average of a sequence of measurements. In other words, the measured quantity is the average $\int_{\Omega} f(x)\varphi(x) \, dx$, and from this reasoning we see that physical variables might be better thought of and treated as such averages rather than as functions $f(x)$ that are precisely defined in every point. This in itself could be viewed as a partial motivation for distribution theory.

Distributions are thus in this sense a generalization of real- or complex-valued functions on Euclidean spaces. While we have stressed the general transition away from pointwise information when dealing with distributions rather than functions, we note that there is in certain cases a way of recovering some of this information from the distribution. If $f$ is continuous at $x_0$, we can take an approximate identity sequence $\{g_k\} \subset C^\infty_c(\Omega)$, and in turn define a sequence $\{\varphi_{k,x_0}\} \subset C^\infty_c(\Omega)$ by putting $\varphi_{k,x_0}(x) = g_k(x - x_0)$. Then it is not difficult to see that $\lim_{k \to \infty} \langle f, \varphi_{k,x_0} \rangle = \lim_{k \to \infty} f * g_k(x_0) = f(x_0)$ by Lemma 2.12, and such a sequence can be found for any point $x_0 \in \Omega$. This highlights that the distribution still retains a great deal of information about the function.

Relating to the discussion leading up to Definition 3.8, this is naturally a consequence of the integral based Definition 3.9 of $\langle f, - \rangle$, and importantly that it takes into
account the behaviour of $f$ on the whole domain. Definition 3.9 further leads to
the fact that two functions giving rise to the same distribution must also be equal
as functions almost everywhere (i.e. can only differ on sets of measure zero). For
example, in $\mathbb{R}$ this implies that there cannot exist any interval of positive length, no
matter how small, where the functions differs on the entire interval. This extends
to $\mathbb{R}^n$ with appropriate adaptions. This is a kind of uniqueness property, which is
highly desirable.

However, it should be stressed that this only applies for functions in $L^1_{loc}$; any such
function gives rise to exactly one distribution and any other locally integrable func-
tions also identifiable with this distribution is equal to $f$ almost everywhere. If $f$
is not locally integrable however, there is typically no way of associating it with just
one distribution in $\mathcal{D}'$. There may be a subset of $C^\infty_c$ on which this is possible, but
there may be infinitely many ways of extending it from there to a distribution on
the whole $C^\infty_c$.

It should be mentioned that some functions not in $L^1_{loc}$ may be still be identified with
distributions in meaningful ways. Consider the typical example of $x^{-m} \not\in L^1_{loc}(\mathbb{R})$
for $m \geq 1$. The problem is that these functions have singularities at the origin at
which they grow too fast to be integrable. For $m = 1$ it is still possible to define the
 corresponding distribution as the Cauchy principal value of the integral in Definition
3.9, i.e. as

$$\langle x^{-1}, \varphi \rangle := \lim_{\epsilon \to 0^+} \int_{-\infty}^{-\epsilon} x^{-1} \varphi(x) \, dx + \int_{\epsilon}^{\infty} x^{-1} \varphi(x) \, dx$$

This is not possible for $m = 2$ though, since $x^{-2}$ is even, so there is no cancellation
and the Cauchy principal value is not applicable. These thus present examples of
functions not in $L^1_{loc}$, one which can be associated with a distribution in a fashion
close to Definition 3.9 and the other which cannot.

However, not all distributions arise from functions. To see that $\mathcal{D}'$ also incorporates
more objects than those of the form in Definition 3.9 we now introduce a distri-
bution that is not a function in the usual sense, the widely recognised Dirac delta
distribution.
Definition 3.11 (Dirac delta distribution). For $\Omega \subset \mathbb{R}^n$ with $y \in \Omega$, we define the Dirac delta distribution in $y$, $\delta(x-y) : C_c^\infty(\Omega) \to \mathbb{C}$, to be

$$\langle \delta(x-y), \varphi \rangle := \varphi(y)$$

For $y = 0$, we write $\delta = \delta(x)$ and call $\delta$ the Dirac delta distribution.

$\delta(x-y) \in \mathcal{D}'(\Omega)$, where linearity is easy to see and continuity follows from the fact that convergence in $C_c^\infty$ of course implies pointwise convergence at any point of $\Omega$, and particularly at $y$. Thus $\varphi_n(y) \to \varphi(y)$ which is precisely the continuity statement for $\delta(x-y)$. $\delta$ is occasionally erroneously referred to as a function (which is also seen in the name) but there is in reality no function corresponding to $\delta$ as a distribution.

To show this requires some measure theory, but we may say that the integral of any such function $f \in L^1_{loc}$ over any compact subset of $\Omega$ not containing 0 would have to be zero. This is because for any such compact subset $S \subset \Omega$ we may construct a function $\varphi \in C_c^\infty$ such that $\text{supp}(\varphi) = S$ and then necessarily $\int_{\text{supp}(\varphi)} f(x)\varphi(x) \, dx = \langle f, \varphi \rangle = 0$ since $\varphi(0) = 0$. We may in fact, for every compact (compactness not strictly required) subset $0 \notin S \subset \Omega$ find a sequence $\varphi_n$ in $C_c^\infty(\Omega)$ such that, pointwise, $\varphi_n \to 1$ on $S$ but 0 otherwise, and with the use of convergence theorems from measure theory one could then show that the integral of $f$ over $S$ must be zero. If one could improve this a little bit to show that this holds for every measurable subset of $\Omega$ not containing 0, this would imply that $f = 0$ almost everywhere on $\Omega$ (see Exercise 2, Chapter 11 in [7]). But according to what was established previously, $f$ and 0 must then define the same distribution and so $\langle f, \varphi \rangle = 0$ for all $\varphi \in C_c^\infty(\Omega)$, and thus $f \neq \delta$ as distributions.

However, $\delta$ as a distribution may be expressed as a limit of distributions corresponding to functions in a way which is very similar to the approximate identities Lemma 2.12. If one considers a sequence of functions $g_k$ which is an approximate identity, one may note that the convolution of $g_k$ with a function $\varphi \in C_c^\infty$ evaluated at 0 is
precisely equal to the value of $\langle g_k, \varphi(-x) \rangle$:

$$\varphi * g_k(0) = \int_{\mathbb{R}^n} \varphi(0 - y) g_k(y) \, dy = \int_{\mathbb{R}^n} g_k(y) \varphi(-y) \, dy = \langle g_k, \varphi(-x) \rangle$$

Taking limits and applying the Lemma 2.12, we obtain $\lim_{k \to \infty} \langle g_k, \varphi(-x) \rangle = \varphi(0) = \lim_{k \to \infty} \langle g_k, \varphi \rangle$ and finally $\lim_{k \to \infty} \langle g_k, \varphi \rangle = \varphi(0) = \langle \delta, \varphi \rangle$ for every $\varphi \in C_c^\infty$. This is an example of a limit of distributions, which quite naturally is defined as follows.

**Definition 3.12** (Limits of distributions). We say that a sequence $\{T_k\} \subset \mathcal{D}'$ converges in $\mathcal{D}'$, or as distributions, to $T \in \mathcal{D}'$ if

$$\lim_{k \to \infty} \langle T_k, \varphi \rangle = \langle T, \varphi \rangle$$

for all $\varphi \in C_c^\infty$.

In contrast to convergence in $C_c^\infty$, convergence as distributions is a weak notion of convergence. For instance, convergence in the $L^1$-norm (on compact subsets) of functions $f_n \in L^1_{loc}$ to a function $f \in L^1_{loc}$ certainly induces convergence of $T_{f_n}$ to $T_f$ as distributions.

To move on we need to introduce a few more basic operations on $\mathcal{D}'$.

**Definition 3.13** (Addition and multiplication with scalars on $\mathcal{D}'$). For $f, g \in \mathcal{D}'$, $\varphi \in C_c^\infty$ and $c \in \mathbb{C}$ we define

(i) $\langle f + g, \varphi \rangle := \langle f, \varphi \rangle + \langle g, \varphi \rangle$

(ii) $\langle cf, \varphi \rangle := c \langle f, \varphi \rangle$

This definition is of course perfectly natural. It is precisely the induced operations that in general makes the dual space a vector space. It is easily shown that linearity and continuity in $\mathcal{D}'$-sense is preserved under these operations (in fact only continuity is needed since linearity is already guaranteed from linear algebra).

We shall now begin with the extension of operations of differential calculus, mainly differentiation, to distributions. Such a generalization must naturally be applicable to functions for which the operation is originally well-defined, and should at least
essentially coincide with the classical definition in those cases. The idea is therefore to start with distributions corresponding to functions for which the operation is already well defined in the classical sense, and determine how the ordinary operation can be carried over to the distributions. This can then be viewed as the restriction of the operation to a subset of distributions, and then one may attempt to extend it to \( \mathcal{D}' \) from there. Thus for some operation \( T \), one considers the distribution corresponding to \( T\varphi \) for functions \( \varphi \) such that \( T\varphi \) is classically defined. In the case of distributions it turns out to be convenient to consider the underlying set of test functions. A general way of extending operations to distributions in this manner is by using adjoint identities.

**Definition 3.14 (Adjoint Identity).** Let \( T \) be an operation on \( C^\infty_c(\Omega) \). An adjoint identity on \( C^\infty_c(\Omega) \) for \( T \) is an equality of the form

\[
\langle T\varphi_1, \varphi_2 \rangle = \int_\Omega T(\varphi_1(x))\varphi_2(x) \, dx = \int_\Omega \varphi_1(x)S(\varphi_2(x)) \, dx = \langle \varphi_1, S\varphi_2 \rangle
\]

for all \( \varphi_1, \varphi_2 \in C^\infty_c(\Omega) \) for some fixed operation \( S \) on \( C^\infty_c(\Omega) \).

An adjoint identity thus enables a shift of the operation \( T \) on the function that is considered as a distribution, to the operation \( S \) on the test function at which the evaluation occurs, and this holds for every \( \varphi_1, \varphi_2 \in C^\infty_c(\Omega) \). One may then proceed to define the operation \( T \) on any distribution \( f \in \mathcal{D}'(\Omega) \) to be \( \langle Tf, \varphi \rangle := \langle f, S\varphi \rangle \). Then the restriction of \( T \) to distributions corresponding to functions on which \( T \) is defined in the original sense, yields consistent results in both interpretations of \( T \), and thus this can be considered a generalization.

Some remarks on this include that there is not in general an adjoint identity for any arbitrary operation \( T \), and one must derive it separately in every case, i.e. there is not a general formula in terms of \( T \). Secondly, we must not always require the original operation \( T \) to map \( C^\infty_c \) into \( C^\infty_c \) (although it perhaps should map \( C^\infty_c \) into \( L^1_{loc} \), so that \( T\varphi \in \mathcal{D}' \), but the other operation \( S \) in an adjoint identity must necessarily preserve the test function space, since otherwise \( \langle Tf, \varphi \rangle = \langle f, S\varphi \rangle \) may not be defined for some \( \varphi \in C^\infty_c \) since for \( f \in \mathcal{D}' \), \( \langle f, \phi \rangle \) is only defined for \( \phi \in C^\infty_c \). Finally, \( S \) must also abide to some sufficient continuity criterion so that the resulting
distribution \( \langle f, S\varphi \rangle \) always is continuous. A sufficient condition would be that \( S \) preserves convergence of test functions, which is to say that \( \varphi_k \to \varphi \) as test functions implies \( S\varphi_k \to S\varphi \) as test functions.

We shall now derive adjoint identities for some basic operations.

**Proposition 3.15** (Adjoint identities for common operations). For \( \varphi_1, \varphi_2 \in C^\infty_c(\Omega) \), \( \psi \in C^\infty(\Omega) \), and, in the case \( \Omega = \mathbb{R}^n \), for fixed \( y \in \mathbb{R}^n \) and translation denoted \( \tau_y \varphi(x) = \varphi(x + y) \), we have that

(i) \( \langle \tau_y \varphi_1, \varphi_2 \rangle = \langle \varphi_1, \tau_{-y} \varphi_2 \rangle \)

(ii) \( \langle \psi \varphi_1, \varphi_2 \rangle = \langle \varphi_1, \psi \varphi_2 \rangle \)

(iii) \( \left\langle \frac{\partial}{\partial x_j} \varphi_1, \varphi_2 \right\rangle = - \left\langle \varphi_1, \frac{\partial}{\partial x_j} \varphi_2 \right\rangle \)

Proof.

(i) \( \langle \tau_y \varphi_1, \varphi_2 \rangle = \int_{\mathbb{R}^n} \varphi_1(x + y) \varphi_2(x) \, dx \)

\[= \int_{\mathbb{R}^n} \varphi_1(x) \varphi_2(x - y) \, dx \]

\[= \langle \varphi_1, \tau_{-y} \varphi_2 \rangle \]

(ii) \( \langle \psi \varphi_1, \varphi_2 \rangle = \int_{\Omega} \psi(x) \varphi_1(x) \varphi_2(x) \, dx \)

\[= \langle \varphi_1, \psi \varphi_2 \rangle \]

(iii) \( \left\langle \frac{\partial}{\partial x_j} \varphi_1, \varphi_2 \right\rangle = \int_{\mathbb{R}^n} \left( \frac{\partial}{\partial x_j} \varphi_1(x) \right) \varphi_2(x) \, dx \)

\[\overset{\text{IBP}}{=} \int_{\mathbb{R}^{n-1}} [\varphi_1(x) \varphi_2(x)]_{x_j=+\infty}^{x_j=-\infty} \, dx - \int_{\Omega} \varphi_1(x) \left( \frac{\partial}{\partial x_j} \varphi_2(x) \right) \, dx \]

\[\overset{(1)}{=} - \int_{\Omega} \varphi_1(x) \left( \frac{\partial}{\partial x_j} \varphi_2(x) \right) \, dx \]

\[= - \left\langle \varphi_1, \frac{\partial}{\partial x_j} \varphi_2 \right\rangle \]

where IBP meaning integration by parts in the \( x_j \)-variable. (1) is true due to the boundary terms vanishing, since \( \varphi_1, \varphi_2 \in C^\infty_c(\Omega) \).
Note that all the above operations map test functions to test functions, by the properties of $C^\infty_c$. Now this leads to the following definitions of these operations on $\mathcal{D}'$.

**Definition 3.16 (Translation, Smooth Multiplication and The Distributional Derivative on $\mathcal{D}'$).** For $f \in \mathcal{D}'(\Omega)$, $\psi \in C^\infty(\Omega)$ and in the case that $\Omega = \mathbb{R}^n$, for fixed $y \in \mathbb{R}^n$, we define

(i) $\langle \tau_y f, \varphi \rangle = \langle f(x+y), \varphi \rangle := \langle f, \varphi(x-y) \rangle = \langle f, \tau_{-y} \varphi \rangle$

(ii) $\langle \psi f, \varphi \rangle := \langle f, \psi \varphi \rangle$

(iii) $\left\langle \frac{\partial}{\partial x_j} f, \varphi \right\rangle := - \left\langle f, \frac{\partial}{\partial x_j} \varphi \right\rangle$

Without question the most interesting and noteworthy out of the above is (iii). It states a generalization of differentiation to a much larger class of objects than just differentiable functions. For instance, it provides meaning to derivatives of objects that are not even functions, such as $\delta$, but it also provides a way to differentiate any locally integrable function as a distribution, and that is infinitely many times. This includes many, in the traditional sense, nowhere-differentiable functions. More connected to the roots of distribution theory is however the fact that this gives meaning to differential equations for a much wider set of objects, including both ordinary functions and strict distributions, so that one may instead consider differential equations in terms of distributions rather than of only differentiable functions.

As one might suspect from the adjoint identity, whenever $f$ is a function, its distributional derivative is consistent with its ordinary derivative in the following sense, under the given conditions.

**Proposition 3.17 (Consistency of Derivatives).** Let $\Omega \subset \mathbb{R}^n$ be open and let $f : \Omega \to \mathbb{R}$. Suppose $f \in L^1_{\text{loc}}(\Omega)$, $\frac{\partial f}{\partial x_j} \in L^1_{\text{loc}}(\Omega)$ and that $\frac{\partial f}{\partial x_j}$ is continuous on $\Omega$ except at a point $y \in \Omega$. If $n = 1$ assume additionally that $f$ is continuous at $y$. Let $T_f$ and $T_{\frac{\partial f}{\partial x_j}}$ denote the distributions corresponding to $f$ and $\frac{\partial f}{\partial x_j}$ respectively. Then

$$\left\langle \frac{\partial}{\partial x_j} T_f, \varphi \right\rangle = \left\langle T_{\frac{\partial f}{\partial x_j}}, \varphi \right\rangle$$

for all $\varphi \in C^\infty_c(\Omega)$. 
Proof. First \( n = 1 \). Putting \( \Omega = (a,b) \), we have

\[
\left\langle \frac{dT}{dx}, \varphi \right\rangle = \int_a^b \frac{df}{dx}(x)\varphi(x) \, dx = - \int_a^y \frac{df}{dx}(x)\varphi(x) \, dx - \int_y^b \frac{df}{dx}(x)\varphi(x) \, dx
\]

\[
= \lim_{t \to y^-} \int_a^t \frac{df}{dx}(x)\varphi(x) \, dx + \lim_{t \to y^+} \int_t^b \frac{df}{dx}(x)\varphi(x) \, dx
\]

\[
= \lim_{t \to y^-} \left( [f(x)\varphi(x)]_a^t - \int_a^t f(x)\frac{d\varphi}{dx}(x) \, dx \right)
\]

\[
+ \lim_{t \to y^+} \left( [f(x)\varphi(x)]_t^b - \int_t^b f(x)\frac{d\varphi}{dx}(x) \, dx \right)
\]

\[
= \lim_{t \to y^-} (f(t)\varphi(t)) + \lim_{t \to y^+} (-f(t)\varphi(t)) - \left( \int_a^y f(x)\frac{d\varphi}{dx}(x) \, dx \right)
\]

\[
= (f(y)\varphi(y) - f(y)\varphi(y)) - \int_a^y f(x)\frac{d\varphi}{dx}(x) \, dx
\]

\[
= - \int_a^b f(x)\frac{d\varphi}{dx}(x) \, dx
\]

\[
= - \left\langle T_f, \frac{d\varphi}{dx} \right\rangle
\]

\[
= \left\langle \frac{d}{dx}T_f, \varphi \right\rangle
\]

Now, for \( n \geq 2 \) the argument is similar, but continuity of \( f \) at \( y \) is not needed because that point does not influence the value of the integral; the boundary terms from partial integration vanishes anyway. For the argument in \( \mathbb{R}^2 \), we refer to [5]. □

Thus if the derivative of a locally integrable function \( f \) is locally integrable and continuous, it corresponds as a distribution to the distributional derivative of \( f \). In higher dimensions, we see that discontinuity of the derivative is allowed up to at least a finite set of points. Note that, why local integrability is required is due to the fact that otherwise, as we noted earlier, \( f \) may not be identifiable with one unique distribution in \( \mathcal{D}' \) and then of course the distributional derivative of \( f \) is not clearly defined for all \( \varphi \in C_c^\infty \).

We now turn to the properties of the distributional derivative, of which most are familiar and inherited from its classical counterpart.
Proposition 3.18 (Basic properties of the distributional derivative). Let \( T, U \in \mathcal{D}'(\Omega) \), \( c_1, c_2 \in \mathbb{R} \) and \( \psi \in C^\infty(\Omega) \). Then

(i) \[ \left\langle \frac{\partial}{\partial x_j} (c_1 T + c_2 U), \varphi \right\rangle = c_1 \left\langle \frac{\partial}{\partial x_j} T, \varphi \right\rangle + c_2 \left\langle \frac{\partial}{\partial x_j} U, \varphi \right\rangle \]

(ii) \[ \left\langle \frac{\partial}{\partial x_j} (\psi T), \varphi \right\rangle = \left\langle \frac{\partial \psi}{\partial x_j} T, \varphi \right\rangle + \left\langle \psi \frac{\partial T}{\partial x_j}, \varphi \right\rangle \]

(iii) \[ \frac{\partial^2}{\partial x_i \partial x_j} T = \frac{\partial}{\partial x_j} \frac{\partial}{\partial x_i} T \text{ for all } i, j \in \{1, \ldots, n\}. \]

(iv) If \( \{T_k\}_{k=1}^\infty \subset \mathcal{D}' \) is such that \( T_k \to T \) in \( \mathcal{D}' \), then \( \frac{\partial^\alpha}{\partial x^\alpha} T_k \to \frac{\partial^\alpha}{\partial x^\alpha} T \) in \( \mathcal{D}' \) for every multi-index \( \alpha \).

Proof.

(i) \[ \left\langle \frac{\partial}{\partial x_j} (c_1 T + c_2 U), \varphi \right\rangle \overset{\text{def.}}{=} - \left\langle (c_1 T + c_2 U), \frac{\partial}{\partial x_j} \varphi \right\rangle 
= -c_1 \left\langle T, \frac{\partial}{\partial x_j} \varphi \right\rangle - c_2 \left\langle U, \frac{\partial}{\partial x_j} \varphi \right\rangle 
= c_1 \left\langle \frac{\partial}{\partial x_j} T, \varphi \right\rangle + c_2 \left\langle \frac{\partial}{\partial x_j} U, \varphi \right\rangle \]

(ii) \[ \left\langle \frac{\partial}{\partial x_j} (\psi T), \varphi \right\rangle = - \left\langle \psi T, \frac{\partial}{\partial x_j} \varphi \right\rangle 
= - \left\langle T, \psi \frac{\partial}{\partial x_j} \varphi \right\rangle 
= - \left\langle T, \psi \frac{\partial}{\partial x_j} \varphi + \frac{\partial \psi}{\partial x_j} \varphi \right\rangle + \left\langle T, \frac{\partial \psi}{\partial x_j} \varphi \right\rangle 
= \left\langle \psi \frac{\partial T}{\partial x_j}, \varphi \right\rangle + \left\langle \frac{\partial \psi}{\partial x_j} T, \varphi \right\rangle \]

(iii) \[ \left\langle \frac{\partial^2}{\partial x_i \partial x_j} T, \varphi \right\rangle = \left\langle T, \frac{\partial^2}{\partial x_j \partial x_i} \varphi \right\rangle = \left\langle T, \frac{\partial^2}{\partial x_i \partial x_j} \varphi \right\rangle = \left\langle \frac{\partial^2}{\partial x_j \partial x_i} T, \varphi \right\rangle \]

where we have used equality of mixed partials for \( \varphi \in C^\infty_c \subset C^2 \).

(iv) We prove the statement for \( \alpha = e_j \), from which the general statement follows inductively. To this end, we have for every \( \varphi \in C^\infty_c \),

\[ \left\langle \frac{\partial}{\partial x_j} T_k, \varphi \right\rangle = - \left\langle T_k, \frac{\partial}{\partial x_j} \varphi \right\rangle \to - \left\langle T, \frac{\partial}{\partial x_j} \varphi \right\rangle = \left\langle \frac{\partial}{\partial x_j} T, \varphi \right\rangle \]

as \( n \to \infty \) since \( \frac{\partial}{\partial x_j} \in C^\infty_c \) and \( T_k \to T \) in \( \mathcal{D}' \), and the proof is complete.
Property (iv) above stands out as it is clearly not inherited from the classical derivative. A sequence of differentiable functions can converge to another function while the sequence of derivatives does not converge at all, or while the limit functions is nowhere differentiable. In this aspect, distributions are much easier to deal with when it comes to limits, and in particular, interchanging limits and differentiation. Now for some particular examples of distributional derivatives.

**Proposition 3.19** (Examples of Distributional derivatives).

(i) \[ \left\langle \frac{\partial}{\partial x_j} \delta, \varphi \right\rangle \overset{\text{def.}}{=} - \left\langle \delta, \frac{\partial}{\partial x_j} \varphi \right\rangle = - \frac{\partial \varphi}{\partial x_j}(0) \]

From this it can be shown that for any multi-index \( \alpha \),
\[ \left\langle \frac{\partial^{\alpha}}{\partial x^{\alpha}} \delta, \varphi \right\rangle = (-1)^{|\alpha|} \left\langle \delta, \frac{\partial^{\alpha} \varphi}{\partial x^{\alpha}} \right\rangle = (-1)^{|\alpha|} \frac{\partial^{\alpha} \varphi}{\partial x^{\alpha}}(0) \]

In general, we then have for any polynomial \( P \), if we let \( \frac{\partial}{\partial x} = \left( \frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_n} \right) \),
\[ \left\langle P \left( \frac{\partial}{\partial x} \right) \delta, \varphi \right\rangle = P \left( - \frac{\partial}{\partial x} \right) \varphi(x) \bigg|_{x=0} \]

(ii) Define the Heaviside function on \( \mathbb{R} \) as \( H(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases} \). Then, as a distribution,
\[ \left\langle \frac{d}{dx} H, \varphi \right\rangle = - \left\langle H, \frac{d}{dx} \varphi \right\rangle \]
\[ = - \int_0^\infty \frac{d\varphi}{dx}(x) \, dx \]
\[ = - \left[ \varphi(x) \right]_0^\infty \]
\[ = \varphi(0) \]
\[ = \left\langle \delta, \varphi \right\rangle \]

Similarly, if one in \( \mathbb{R}^n \) defines \( H_n \) as
\[ H_n(x) = H(x_1)H(x_2)\ldots H(x_n) = \begin{cases} 1 & x_j \geq 0, \forall j : 1 \leq j \leq n \\ 0 & \text{otherwise} \end{cases} \]

then
\[\langle \frac{\partial}{\partial x_1} \ldots \frac{\partial}{\partial x_n} H, \varphi \rangle = (-1)^n \langle H, \frac{\partial}{\partial x_1} \ldots \frac{\partial}{\partial x_n} \varphi \rangle \]
\[= (-1)^n \int_0^\infty \cdots \int_0^\infty \frac{\partial}{\partial x_1} \ldots \frac{\partial}{\partial x_n} \varphi(x) \, dx_1 \ldots dx_n \]
\[IBP \quad (-1)^n \int_0^\infty \cdots \int_0^\infty (-1) \frac{\partial}{\partial x_1} \ldots \frac{\partial}{\partial x_n} \varphi(x) \bigg|_{x_1=0} \, dx_2 \ldots dx_n = \ldots \]
\[= (-1)^n \int_0^\infty \cdots \int_0^\infty (-1)^{k-1} \frac{\partial}{\partial x_k} \ldots \frac{\partial}{\partial x_n} \varphi(x) \bigg|_{x_1,\ldots,x_{k-1}=0} \, dx_k \ldots dx_n \]
\[= (-1)^n (-1)^n \varphi(0) \]
\[= \varphi(0) \]
\[= \langle \delta, \varphi \rangle \]

(iii) \[\frac{d}{dx} |x|, \varphi \rangle = - \langle |x|, \frac{d\varphi}{dx} \rangle \]
\[= \int_{-\infty}^0 x \frac{d\varphi}{dx} (x) \, dx - \int_0^\infty x \frac{d\varphi}{dx} (x) \, dx \]
\[IBP \quad [x \varphi(x)]^0_{-\infty} - \int_{-\infty}^0 \varphi (x) \, dx - [x \varphi(x)]_0^\infty + \int_0^\infty \varphi (x) \, dx \]
\[= \langle H - IH, \varphi \rangle \]

At this point we have seen several purposes of the test functions, which justify and make sense of the harsh criteria imposed on the test function space. First of all, they control the behaviour of the distributions through their compact support property, especially allowing a rather general class of functions to be identified as distribution, as discussed earlier. Secondly, by letting the test functions have desirable and strong properties, for instance smoothness, and in conjunction with the test function space being preserved under the operation corresponding to the properties, e.g. differentiation, it is possible to lift the properties of the test functions to the level of distributions with the use of adjoint identities. As we have now seen, this enables us to generalize differentiation to general distributions, which all are differentiable infinitely many times in this more general sense. This allows us to view differential equations in terms of distributions. The next step is to extend the Fourier transform to distributions in a similar fashion - the Fourier transform is an indispensable tool for dealing with linear partial differential equations, and it is naturally useful
if this could be extended to distributions as well. However, it turns out that this
is not possible to achieve for general distributions - at least not in such a way that
their Fourier transforms are also general distributions - and the reason is that \( C_\infty \) is
not preserved under \( \mathcal{F} \). This is a consequence of a Payley-Weiner(-Schwartz) type
theorem for \( C_\infty \) asserting that the Fourier transform of a \( C_\infty \) function is extendable
to an complex analytic function, and thus cannot have compact support unless it
is zero everywhere \([3,5]\). The situation could therefore even be described as the
opposite of preservation: \( \varphi \in C_\infty \) and \( \mathcal{F}\varphi \in C_\infty \) implies \( \varphi = 0 \) identically \([3,5]\). On a short note, it is strictly speaking possible to construct the Fourier transform
of general distributions but this is technical and introduces another class of objects
called ultradistributions \([3]\).

The primary path to extending the Fourier transform, which we will pursue here, is
to consider a subclass of distributions. Here the full importance of \( \mathcal{S} \) will finally be
realized. The functions in \( \mathcal{S} \) have many strong properties in common with those in
\( C_\infty \), primarily smoothness and integrability, but the important differences are that
\( \mathcal{S} \) has a weaker support condition but on the other hand is preserved under \( \mathcal{F} \), by
Theorem 2.21. The course of action is therefore to consider distributions defined
on the test function space \( \mathcal{S} \) rather than \( C_\infty \) - this yields the subclass of tempered
distributions. For \( \mathcal{S} \) to act as a test functions space, it needs a notion of convergence
which is defined slightly differently from that of \( C_\infty \).

**Definition 3.20** (Convergence in \( \mathcal{S} \)). Let \( \{ \varphi_k \} \) be a sequence in \( \mathcal{S} \). We say that
\( \varphi_k \) converges to \( \varphi \) in \( \mathcal{S} \) or in \( \mathcal{S} \)-sense, written

\[
\varphi_k \to \varphi \quad \text{or} \quad \lim_{k \to \infty} \varphi_k = \varphi
\]

if for every integer \( m \geq 0 \) and multi-index \( \alpha \),

\[
\lim_{k \to \infty} \left\| |x|^m \frac{\partial^n \varphi_k}{\partial x^\alpha} - |x|^m \frac{\partial^n \varphi}{\partial x^\alpha} \right\|_\infty = 0
\]

**Definition 3.21** (Tempered Distributions, \( \mathcal{S}'(\mathbb{R}^n) \)). A tempered distribution on \( \mathbb{R}^n \)
is a linear and continuous map \( \langle f, - \rangle : \mathcal{S}(\mathbb{R}^n) \to \mathbb{R}, \varphi \mapsto \langle f, \varphi \rangle \). That is, \( \langle f, \varphi \rangle \) satisfies
(1) (Linearity) \( \langle f, c_1 \varphi_1 + c_2 \varphi_2 \rangle = c_1 \langle f, \varphi_2 \rangle + c_2 \langle f, \varphi_2 \rangle \)
for all \( \varphi_1, \varphi_2 \in \mathcal{S}(\mathbb{R}^n) \) and \( c_1, c_2 \in \mathbb{R} \).

(2) (Continuity) \( \lim_{k \to \infty} \langle f, \varphi_k \rangle = \langle f, \varphi \rangle \) whenever \( \varphi_k \to \varphi \) in \( \mathcal{S} \)-sense.

The set of all tempered distributions on \( \mathbb{R}^n \) is denoted \( \mathcal{S}'(\mathbb{R}^n) \) or simply \( \mathcal{S} \) when \( n \) is obvious or irrelevant in the context.

By recalling that \( C_c^\infty \subset \mathcal{S} \), every distribution in \( \mathcal{S}' \) must also be defined on \( C_c^\infty \). One can show that the continuity in \( \mathcal{S}' \) on \( \mathcal{S} \) implies continuity in \( \mathcal{D}' \)-sense on \( C_c^\infty \). Thus every distribution in \( \mathcal{S}' \) also defines a distribution in \( \mathcal{D}' \), i.e. we have the reverse inclusion \( \mathcal{S}' \subset \mathcal{D}' \). However, if we consider a distribution \( T_f \) arising from a function \( f \), a sufficient criterion for \( T_f \in \mathcal{D}' \) is \( f \in L^1_{loc} \). But this allows \( f \) to be unbounded and grow arbitrarily fast at infinity, all because of the compact support property of test functions in \( C_c^\infty \) which enables the \( m \) to control this by being identically zero outside a bounded region. However, this is not enough for \( T_f \in \mathcal{S}' \) because the functions in \( \mathcal{S} \) are in general only capable of controlling functions of at most polynomial growth at infinity, and if \( f \) grows faster than this then it may happen that \( |\langle T_f, \varphi \rangle| = +\infty \) for some \( \varphi \in \mathcal{S} \), so that \( T_f \) cannot be sensibly defined on \( \mathcal{S} \). Thus there are distributions in \( \mathcal{D}' \) which are not tempered distributions and the previous inclusion is strict - one example is the exponential function \( e^{\|x\|} \). To summarize, when extending the test functions space from \( C_c^\infty \) to \( \mathcal{S} \), we ease the harshness of the support condition on the test functions. This, however, translates to slightly harsher requirements on the tempered distributions, which cannot behave quite as wildly as the general distributions since the controlling properties of the test functions in \( \mathcal{S} \) are weaker. In return, as we will see, this allows us to extend the Fourier transform on \( \mathcal{S}' \) in a satisfactory fashion.

We will relate functions to distributions in \( \mathcal{S}' \) in precisely the same way as in Definition 3.7 for \( \mathcal{D}' \), but the above discussion indicates that we need to require something stronger than local integrability. The following criterion is sufficient.
Definition 3.22 \((L^1_{\text{pol}}\text{ and functions of slow growth})\). If a function \(f \in L^1_{\text{loc}}(\mathbb{R}^n)\) satisfies the estimate
\[
\int_{|x| \leq R} |f(x)| \, dx \leq MR^k
\]
whenever \(R \geq R_0\) for some constants \(R_0, M \in \mathbb{R}, k \in \mathbb{N}\), then we say that \(f \in L^1_{\text{pol}}\).

If a function \(g\) satisfies
\[
|g(x)| \leq C \left(1 + |x|^k\right), \quad \forall x \in \mathbb{R}^n
\]
for some constants \(C \in \mathbb{R}, k \in \mathbb{N}\), then we say that \(g\) is a function of slow growth or slowly growing.

From now on, when we speak of operations on \(S'\) that we previously defined for \(D'\) we will mean the same operation except changing \(L^1_{\text{loc}}\) to \(L^1_{\text{pol}}\), \(D\) to \(S\) and \(D'\) to \(S'\) in the definitions. This will be the case for addition, multiplication by scalars, translation, differentiation and adjoint identities. However, in the case of multiplication by \(C^\infty\)-functions we need to require the function and all of its derivatives to be slowly growing - otherwise the multiplication with the function will not preserve \(S\) and the way of defining multiplication with an adjoint identity as in the \(D'\) case will fail.

Now we derive the adjoint identities for the Fourier transform and convolution on \(S\).

**Proposition 3.23 (Adjoint identities of convolution, \(I, \mathcal{F}\) and \(\mathcal{F}^{-1}\) on \(S\)).** For all \(\varphi_1, \varphi_2, \varphi_3 \in S\) the following holds

(i) \(\langle I\varphi_1, \varphi_2 \rangle = \langle \varphi_1, I\varphi_2 \rangle\)

(ii) \(\langle \varphi_1 * \varphi_2, \varphi_3 \rangle = \langle \varphi_2, I(\varphi_1) \ast \varphi_3 \rangle = \langle \varphi_1, I(\varphi_2) \ast \varphi_3 \rangle\)

(iii) \(\langle \mathcal{F}\varphi_1, \varphi_2 \rangle = \langle \varphi_1, \mathcal{F}\varphi_2 \rangle\)

(iv) \(\langle \mathcal{F}^{-1}\varphi_1, \varphi_2 \rangle = \langle \varphi_1, \mathcal{F}^{-1}\varphi_2 \rangle\)

**Proof.**

(i) \(\langle I\varphi_1, \varphi_2 \rangle = \int_{\mathbb{R}^n} \varphi_1(-x)\varphi_2(x) \, dx = \int_{\mathbb{R}^n} \varphi_1(x)\varphi_2(-x) \, dx = \langle \varphi_1, I\varphi_2 \rangle\)
\[ \langle \varphi_1 \ast \varphi_2, \varphi_3 \rangle = \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \varphi_1(x - y) \varphi_2(y) \varphi_3(x) \, dy \, dx \]

\[ \overset{(1)}{= \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \varphi_1(x - y) \varphi_2(y) \varphi_3(x) \, dx \, dy} \]

\[ = \int_{\mathbb{R}^n} \varphi_2(y) \left( \int_{\mathbb{R}^n} \varphi_1(-(y - x)) \varphi_3(x) \, dx \right) \, dy \]

\[ = \int_{\mathbb{R}^n} \varphi_2(y) ((I \varphi_1) \ast \varphi_3(y)) \, dy \]

\[ = \langle \varphi_2, (I \varphi_1) \ast \varphi_3 \rangle \]

\[ = \langle \varphi_1, (I \varphi_2) \ast \varphi_3 \rangle \]

The change of order of integration in (1) is permitted since the integrals on either side are absolutely convergent, due to the fact that \( \varphi_1, \varphi_2, \varphi_3 \in \mathcal{S}, \mathcal{S} \) is an algebra and any function in \( \mathcal{S} \) is absolutely integrable. The last equality follows from commutativity of \( \ast \).

(iii) This follows directly from an application of Proposition 2.22,

\[ \langle \mathcal{F} \varphi_1, \varphi_2 \rangle = \int_{\mathbb{R}^n} \mathcal{F} \varphi_1(\xi) \varphi_2(\xi) \, d\xi = \int_{\mathbb{R}^n} \varphi_1(x) \mathcal{F} \varphi_2(x) \, dx = \langle \varphi_1, \mathcal{F} \varphi_2 \rangle \]

(iv) \[ \langle \mathcal{F}^{-1} \varphi_1, \varphi_2 \rangle = (2\pi)^{-n} \langle (\mathcal{F} \varphi_1), \varphi_2 \rangle \]

\[ = (2\pi)^{-n} \langle \varphi_1, (I \varphi_2) \rangle \]

\[ = \langle \varphi_1, (2\pi)^{-n} I (F \varphi_2) \rangle \]

\[ = \langle \varphi_1, \mathcal{F}^{-1} \varphi_2 \rangle \]

\[ \square \]

Similar to before, this motivates the definition of these operations on tempered distributions in a straight-forward manner, as follows.

**Definition 3.24** (Inversion, convolution and Fourier Transforms on \( \mathcal{S}' \)). For \( f \in \mathcal{S}' \) and \( \varphi_1, \varphi_2 \in \mathcal{S} \) we define

(i) \( \langle I f, \varphi_1 \rangle := \langle f, I \varphi_1 \rangle \)

(ii) \( \langle \varphi_1 \ast f, \varphi_2 \rangle = \langle f \ast \varphi_1, \varphi_2 \rangle := \langle f, (I \varphi_1) \ast \varphi_2 \rangle \)

(iii) \( \langle \mathcal{F} f, \varphi_1 \rangle = \langle f, \mathcal{F} \varphi_1 \rangle \)

(iv) \( \langle \mathcal{F}^{-1} f, \varphi_1 \rangle = \langle f, \mathcal{F}^{-1} \varphi_1 \rangle \)
Note that, as in the case for multiplication, convolution is only defined between a function and a distribution and not between arbitrary distributions.

The extension of the Fourier transform to tempered distributions is noteworthy in the same regard as the distributional derivative. Earlier, in the first section when the Fourier transform of functions were defined, one typically must require some kind of integrability, which implicitly means decay at infinity in a certain sense, in order to ensure existence of the Fourier transform. However, functions identifiable as tempered distributions may be unbounded as they are allowed to grow polynomially. Thus, this enables Fourier transforms of a much larger, less restricted class of functions, including for instance all polynomials, which are not integrable over $\mathbb{R}^n$ in any case except for the 0 polynomial.

**Proposition 3.25** (Properties of $\ast$, $\mathcal{F}$ and $\mathcal{F}^{-1}$ on $\mathcal{S}'$).

If $f \in \mathcal{S}'$ and $\varphi_1 \in \mathcal{S}$ then

(i) $\frac{\partial}{\partial x_j} (f \ast \varphi_1) = \frac{\partial f}{\partial x_j} \ast \varphi_1 = \frac{\partial \varphi_1}{\partial x_j} \ast f$

(ii) $\mathcal{F}^{-1} \mathcal{F} f = f = \mathcal{F} \mathcal{F}^{-1} f$ *(Fourier Inversion Theorem on $\mathcal{S}'$)*

(iii) $\mathcal{F}^{-1} f = (2\pi)^{-n} I F f$

(iv) $\mathcal{F} (f \ast \varphi_1) = \mathcal{F} f \mathcal{F} \varphi_1$

(v) $\mathcal{F} (f \varphi_1) = (2\pi)^{-n} \mathcal{F} f \ast \mathcal{F} \varphi_1$

(vi) $\mathcal{F}^{-1} (f \ast \varphi_1) = (2\pi)^{-n} \mathcal{F}^{-1} f \mathcal{F}^{-1} \varphi_1$

(vii) $\mathcal{F}^{-1} (f \varphi_1) = \mathcal{F}^{-1} f \ast \mathcal{F}^{-1} \varphi_1$

(viii) $\mathcal{F} \left( \frac{\partial}{\partial x_j} f \right) = -i x_j \mathcal{F} f$

(ix) $\mathcal{F} (x_j f) = -i \frac{\partial}{\partial x_j} \mathcal{F} f$

(x) $\mathcal{F}^{-1} \left( \frac{\partial}{\partial x_j} f \right) = i x_j \mathcal{F}^{-1} f$

(xi) $\mathcal{F}^{-1} (x_j f) = i \frac{\partial}{\partial x_j} \mathcal{F}^{-1} f$
Proof.

(i) \[ \left\langle \frac{\partial}{\partial x_j} (f \ast \phi), \varphi_2 \right\rangle = - \left\langle f \ast \varphi_1, \frac{\partial \varphi_2}{\partial x_j} \right\rangle \]
\[= - \left\langle f, I(\varphi_1) \ast \frac{\partial \varphi_2}{\partial x_j} \right\rangle \]
\[= - \left\langle f, \frac{\partial}{\partial x_j} (I(\varphi_1) \ast \varphi_2) \right\rangle \]
\[= \left\langle \frac{\partial f}{\partial x_j}, I(\varphi_1) \ast \varphi_2 \right\rangle \]
\[= \left\langle \frac{\partial f}{\partial x_j} \ast \varphi_1, \varphi_2 \right\rangle \]
where we have used the definitions of the distributional derivative and convolution, and properties of convolutions of functions. To show the second equality, we may continue from the expression on the second row above in the following way, noting that \( \frac{\partial}{\partial x_j} I \varphi = \frac{\partial}{\partial x_j} \varphi(-x) = -I \left( \frac{\partial \varphi}{\partial x_j} \right) \) and that constant factors in front of one function may be taken out of the convolution.

\[- \left\langle f, I(\varphi_1) \ast \frac{\partial \varphi_2}{\partial x_j} \right\rangle = - \left\langle f, \left( \frac{\partial (I \varphi_1)}{\partial x_j} \right) \ast \varphi_2 \right\rangle \]
\[= - \left\langle f, - I \left( \frac{\partial \varphi_1}{\partial x_j} \right) \ast \varphi_2 \right\rangle \]
\[= - \left\langle f, I \left( \frac{\partial \varphi_1}{\partial x_j} \right) \ast \varphi_2 \right\rangle \]
\[= \left\langle \frac{\partial \varphi_1}{\partial x_j} \ast f, \varphi_2 \right\rangle \]

(ii) Due to the work done in Proposition 3.23 with the adjoint identities, we have due to the definition of \( \mathcal{F} \) and \( \mathcal{F}^{-1} \) for distributions and for functions that
\[ \left\langle \mathcal{F}^{-1} \mathcal{F} f, \varphi_1 \right\rangle = \left\langle \mathcal{F} f, \mathcal{F}^{-1} \varphi_1 \right\rangle \]
\[= \left\langle f, \mathcal{F} \mathcal{F}^{-1} \varphi_1 \right\rangle \]
\[= \left\langle f, \varphi_1 \right\rangle \]
\[= \left\langle \mathcal{F} \mathcal{F}^{-1} f, \varphi_1 \right\rangle \]
\[= \left\langle \mathcal{F}^{-1} f, \varphi_1 \right\rangle \]
where FIT refers to the Fourier Inversion theorem, Corollary 2.16.

(iii) $\langle \mathcal{F}^{-1} f, \varphi_1 \rangle = \langle f, \mathcal{F}^{-1} \varphi_1 \rangle$

$= \langle f, (2\pi)^{-n} I \mathcal{F} \varphi_1 \rangle$

$= \langle f, (2\pi)^{-n} \mathcal{F} I \varphi_1 \rangle$

$= (2\pi)^{-n} \langle \mathcal{F} f, I \varphi_1 \rangle$

$= (2\pi)^{-n} \langle I \mathcal{F} f, \varphi_1 \rangle$

(iv) $\langle \mathcal{F} (f \ast \varphi_1), \varphi_2 \rangle = \langle f \ast \varphi_1, \mathcal{F} \varphi_2 \rangle$

$= \langle f, \mathcal{F} \varphi_1 \rangle$

$\stackrel{\text{FIT}}{=} \langle f, \mathcal{F} \mathcal{F}^{-1} (I(\varphi_1) \ast \mathcal{F} \varphi_2) \rangle$

$= \langle \mathcal{F} f, \mathcal{F}^{-1} (I(\varphi_1) \ast \mathcal{F} \varphi_2) \rangle$

$\stackrel{\text{Prop. 2.17}}{=} \langle \mathcal{F} f, (2\pi)^n \mathcal{F}^{-1} I(\varphi_1) \ast \mathcal{F}^{-1} \mathcal{F} \varphi_2 \rangle$

$\stackrel{\text{FIT}}{=} \langle \mathcal{F} f, \mathcal{F} \varphi_1 \ast \varphi_2 \rangle$

$= \langle \mathcal{F} f \mathcal{F} \varphi_1, \varphi_2 \rangle$

where again FIT refers to Corollary 2.16. We have also used that $(2\pi)^n \mathcal{F}^{-1} I = \mathcal{F}$ on $S$.

(v) $\langle \mathcal{F}(f \varphi_1), \varphi_2 \rangle = \langle f \varphi_1, \mathcal{F} \varphi_2 \rangle$

$= \langle f, \varphi_1 \mathcal{F} \varphi_2 \rangle$

$\stackrel{\text{FIT}}{=} \langle f, \mathcal{F} \mathcal{F}^{-1} (\varphi_1 \mathcal{F} \varphi_2) \rangle$

$= \langle \mathcal{F} f, \mathcal{F}^{-1} (\varphi_1 \mathcal{F} \varphi_2) \rangle$

$\stackrel{\text{Prop. 2.17}}{=} \langle \mathcal{F} f, \mathcal{F}^{-1} \varphi_1 \ast \mathcal{F}^{-1} \mathcal{F} \varphi_2 \rangle$

$\stackrel{\text{FIT}}{=} \langle \mathcal{F} f, (2\pi)^{-n} I \mathcal{F} \varphi_1 \ast \varphi_2 \rangle$

$= (2\pi)^{-n} \langle \mathcal{F} f \ast \mathcal{F} \varphi_1, \varphi_2 \rangle$

(vi) This follows by first applying (v) to the distribution $\mathcal{F}^{-1} f$ (in place of $f$) and the test function $\mathcal{F}^{-1} \varphi_1$ (in place of $\varphi_1$), which after simplification with the Fourier inversion theorem and (i) yields:
\[ \langle \mathcal{F} (\mathcal{F}^{-1} f \mathcal{F}^{-1} \varphi_1), \varphi_2 \rangle = (2\pi)^{-n} \langle \mathcal{F} \mathcal{F}^{-1} f * \mathcal{F} \mathcal{F}^{-1} \varphi_1, \varphi_2 \rangle \]

\[ \overset{\text{FIT}}{=} (2\pi)^{-n} \langle f * \varphi_1, \varphi_2 \rangle \]

and then the statement follows by applying \( \mathcal{F}^{-1} \) and using (i),

\[ \langle \mathcal{F}^{-1} \mathcal{F} (\mathcal{F}^{-1} f \mathcal{F}^{-1} \varphi_1), \varphi_2 \rangle = (2\pi)^{-n} \langle \mathcal{F}^{-1} f \mathcal{F}^{-1} \varphi_1, \varphi_2 \rangle \]

\[ = (2\pi)^{-n} \langle \mathcal{F}^{-1} (f * \varphi_1), \varphi_2 \rangle \]

(vii) The proof is essentially analogous to (vi), but this follows instead from (iv). Apply (iv) to the distribution \( \mathcal{F}^{-1} f \) convoluted with the test function \( \mathcal{F}^{-1} \varphi_1 \), and apply (i) to obtain the equality

\[ \langle \mathcal{F} (\mathcal{F}^{-1} f * \mathcal{F}^{-1} \varphi_1), \varphi_2 \rangle \overset{(i)}{=} \langle \mathcal{F} \mathcal{F}^{-1} f * \mathcal{F} \mathcal{F}^{-1} \varphi_1, \varphi_2 \rangle \]

\[ = \langle f \varphi_1, \varphi_2 \rangle \]

and the application of \( \mathcal{F}^{-1} \) to both sides and then simplification using (i) again yields the desired result:

\[ \langle \mathcal{F}^{-1} \mathcal{F} (\mathcal{F}^{-1} f * \mathcal{F}^{-1} \varphi_1), \varphi_2 \rangle \overset{(i)}{=} \langle \mathcal{F}^{-1} f * \mathcal{F}^{-1} \varphi_1, \varphi_2 \rangle \]

\[ = \langle \mathcal{F}^{-1} (f \varphi_1), \varphi_2 \rangle \]

\[ \square \]

Convolution between a tempered distribution and a smooth function of slow growth may however be viewed slightly differently. It is actually possible to not only define this as a tempered distribution, but as a proper function. It is done in the following way.

**Definition 3.26 (Convolution on \( \mathcal{S}' \) as a function).** For \( f \in \mathcal{S}' \) and \( \varphi \in \mathcal{S} \) we define

\[ \varphi * f(x) = \langle f, \tau_x (I \varphi) \rangle (x) = \langle f, \varphi(x - y) \rangle (x) \]

where \( x \in \mathbb{R}^n \) is the independent variable.
This is motivated by the fact that if a function \( f \in L^1_{\text{pol}} \) defines a tempered distribution, we have that for any \( x \in \mathbb{R}^n \),
\[
\varphi \ast f(x) = \int_{\mathbb{R}^n} \varphi(x - y) f(y) \, dy = \langle f, \varphi(x - y) \rangle = \langle f, \tau_x \varphi(-y) \rangle = \langle f, \tau_x (I \varphi) \rangle
\]
and this is defined for every \( x \), and thus one can take the right-hand side as a definition for all tempered distributions since it is defined for any \( f \in \mathcal{S}' \). One may show that the function defined in this way is smooth, i.e. \( \langle T, \tau_x (I \varphi) \rangle \in C^\infty \), and even of slow growth, for any tempered distribution \( T \) and \( \varphi \in \mathcal{S} \). Thus, this function is identifiable with a tempered distribution. Moreover, this interpretation of the convolution is consistent with that in Definition 3.24 - that is, the distribution \( T_{\varphi \ast f} \) associated to the function \( \varphi \ast f \) in fact agrees with the distributional convolution defined through the adjoint identity. We state this without proof below. To prove this is rather technical, and can be achieved for instance by showing that for any \( \phi \in \mathcal{S} \) one can write a sequence of Riemann sums converging to \( \varphi \ast \phi \) in \( \mathcal{S} \), and then leverage the distributional continuity of \( f \). For such a proof for \( \mathcal{D}' \) and \( C^\infty_{c} \) in the case \( \Omega = \mathbb{R} \), see [6]. For a more technical discussion we refer to [3].

**Proposition 3.27** (Equality of convolution). Let \( f \in \mathcal{S}' \) and \( \varphi \in \mathcal{S} \). Then it holds that
\[
T_{\varphi \ast f(x)} = \varphi \ast f
\]

In general the treatment of convolutions and products in distribution theory is considerably involved. It requires the notion of support of distributions, and typically proceeds through the development of tensor products of distributions. It is interesting and unfortunate that while distribution theory allows powerful generalizations of differentiation and the Fourier transform - which classically in some sense are rather restricted operations - it does not permit the same for multiplication and convolution. It turns out that a fully general notion of multiplication or convolution between two distributions is not achievable for all distributions [6]. However, a more general approach to convolutions between distributions can be found in [3, 9], and a more extensive treatment of multiplication and convolution is given in [6].
We now turn to some concrete examples. It is at this point easy to see that \( \delta \in \mathcal{S}' \), and thus we may consider its Fourier transform and convolution. These are rather important and occur quite frequently in various contexts, and distribution theory gives them rigorous meaning.

**Proposition 3.28** (Interpretations of the Dirac delta distribution). \( \delta \in \mathcal{S}' \) and for \( \varphi \in \mathcal{S} \) the following is true,

(i) \( \varphi \star \delta(x) = \varphi(x) \)

(ii) \( \mathcal{F}\delta = 1 \)

**Proof.**

(i) \( \varphi \star \delta(x) = \langle \delta(y), \tau_x \varphi(-y) \rangle = \tau_x \varphi(0) = \varphi(x) \)

(ii) \( \langle \mathcal{F}\delta, \varphi \rangle = \langle \delta, \mathcal{F}\varphi \rangle = \mathcal{F}\varphi(0) = \int_{\mathbb{R}^n} \varphi(x) \, dx = \langle 1, \varphi \rangle \)

Part (i) of Proposition \( \text{3.28} \) holds, of course, both interpreted as a function and in the sense of distributions. It is the foundation of the concept of a fundamental solution (or Green’s functions) to a partial differential operator. Essentially, for a linear partial differential operator \( D = \sum_{\alpha} c_{\alpha} \frac{\partial^{\alpha}}{\partial x^\alpha} \) the fundamental solution is a distribution \( T \) satisfying

\[
DT = \delta
\]

The point is that if \( T \) is known, then (i) in Proposition \( \text{3.28} \) yields the solution \( U \) for

\[
DU = f(x)
\]

simply as \( T \star f \), since

\[
D (T \star f) = DT \star f = \delta \star f = f(x)
\]

From the theory we have developed here, this would work for \( f \in \mathcal{S} \).
The second part of Proposition 3.28 on the other hand provides a valid interpretation of
\[
\delta(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ix\xi} \, d\xi
\]

We conclude this section with a few more examples of Fourier transforms of distributions.

**Proposition 3.29** (Examples of distributional Fourier transforms in \(S'(\mathbb{R})\)).

In \(S'(\mathbb{R})\), where \(\frac{1}{x}\) is the Cauchy principal value distribution of the function \(\frac{1}{x}\) and with \(\frac{1}{x^2} = -\frac{d}{dx} \frac{1}{x}\), we have:

(i) \(\mathcal{F}x^k = 2\pi (-i)^k \delta^{(k)}(x)\) for \(k \geq 0\),

(ii) \(\mathcal{F}p(x) = \sum_{k=0}^{N} c_k 2\pi (-i)^k \delta^{(k)}(x)\) for a polynomial \(p(x) = \sum_{k=0}^{N} c_k x^k\),

(iii) \(\mathcal{F}H(x) = \frac{i}{x} + \pi \delta(x)\),

(iv) \(\mathcal{F} \text{sgn}(x) = \frac{2i}{x}\),

(v) \(\mathcal{F} \frac{1}{x} = -2\pi iH(-x) + \pi i = \pi i \text{sgn}(x)\),

(vi) \(\mathcal{F}|x| = -\frac{2}{x^2}\)

**Proof.** I will show (i) for the case \(k = 1\), and then (i) follows easily by iterating Proposition 3.25 (ix), and in turn (ii) follows by linearity. Similarly, I will show (iii), from which (iv)-(vi) follows with some additional applications of various parts of Proposition 3.25 and the relations \(\text{sgn}(x) = H(x) - H(-x) = 1 - 2H(-x)\) and \(|x| = x \text{sgn}(x)\).

(i) \(\mathcal{F}x = \mathcal{F}(x \cdot 1) = -i \frac{d}{dx} \mathcal{F}1 = 2\pi (-i) \frac{d}{dx} (1 \mathcal{F}^{-1}1) = 2\pi (-i) \delta'(x)\)

where we have used Propositions 3.25 and 3.28 and the fact that \(I \delta = \delta\).

(iii) To show this we use the following lemma, which we take as granted without proof (see Exercise 2.3 in [9]):

**Lemma 3.30.** If \(xu' + u = 0\) for \(u \in S'(\mathbb{R})\), then \(u = \frac{A}{x} + B\delta\) for some constants \(A, B \in \mathbb{C}\).
We observe that

\[ -ix\mathcal{F}H(x) = \mathcal{F}\left(\frac{d}{dx} H(x)\right) = \mathcal{F}\delta = 1 \]

Thus, multiplying with \(i\) and differentiating both sides we obtain

\[ x\left(\frac{d}{dx}\mathcal{F}H(x)\right) + \mathcal{F}H(x) = 0 \]

By the lemma, we can then conclude that \(\mathcal{F}H(x) = \frac{A}{x} + B\delta\). It remains to determine \(A\) and \(B\). These are determined by the following system of equations.

\[ -ix\mathcal{F}H(x) = 1 \iff -i\left(\frac{A}{x} + B\delta(x)\right) = 1 \]

\[ \mathcal{F}(H(x) + H(-x)) = \mathcal{F}1 \iff \frac{A}{x} + B\delta + \frac{A}{-x} + B\delta(-x) = 2\pi\delta(x) \]

Using the facts that \(x\frac{1}{x^n} = \frac{1}{x^{n-1}}\) for \(n = 1, 2\), \(x\delta = 0\), \(I\delta = \delta\) and \(\frac{A}{-x} = -\frac{A}{x}\) these simplify to

\[ -iA = 1 \]

\[ 2B\delta(x) = 2\pi\delta(x) \]

which implies \(A = i\) and \(B = \pi\), proving (iii).

\[\square\]

As a last remark of this section, we note that none of the functions corresponding to the distributions above are in \(L^1(\mathbb{R})\) or \(L^2(\mathbb{R})\) - and they lack Fourier transforms in the classical sense.

4. Partial Differential Equations

We will now focus on illustrating and solving a few problems dealing with linear partial differential equations (PDE:s) using the theory developed previously. The main tools for this purpose is of course the Fourier transform and related results. The power of distribution theory on the other hand does not so much introduce new tools or enter explicitly in the solution of PDE:s, but rather remarkably enhances the applicability of the existing methods, i.e. the Fourier transform. Moreover, it
also vastly increases the realm of solutions that are considered, as we can view the equations in terms of distributions. Essentially, distribution theory frees us from many otherwise necessary preemptive assumptions on the solutions of the PDE:s, and allows a more carefree treatment. One must note however, that some conditions are still imposed, in particular if the solution is a function. If we use the Fourier transform methods, the obtained solutions are distributions in $S'$, so if they are functions they must lie in $L^1_{pol}$ and thus cannot grow arbitrarily fast at infinity. One interesting example concerns harmonic functions, which are solutions to Laplace’s equation

$$\nabla^2 f = 0$$

There are indeed harmonic functions growing faster than any polynomial - and these cannot be retrieved by employing the Fourier transform even in the distributional sense [5]. This is an important aspect to keep in mind. Nevertheless, distribution theory obviously drastically widens the scope of solutions that can be found through the Fourier transform, compared to the classical case.

Before we begin, we define the concepts of weak and classical solutions.

**Definition 4.1** (Weak and classical solutions). A function or distribution satisfying a partial differential equation in the distributional sense is called a weak solution. A function which is continuously differentiable that satisfies a partial differential equation is called a classical solution.

Due to the consistency theorem for derivatives, it is clear that classical solutions are also weak solutions. Whenever one solves a PDE in the distributional sense and obtains a weak solution, one may then observe if the weak solution corresponds to a locally integrable and continuously differentiable function. If this is the case, then by consistency the function must also be a classical solution. In this way distribution theory is not merely useful for finding weak solutions, but may also help in finding classical solutions. There are more sophisticated ways of determining classical solutions (or at least solutions that are proper functions) from distribution
solutions - granted this is possible - though this is the topic of Sobolev theory and lies outside of our scope. Due to this, the examples we will consider here will mostly exemplify the use of the Fourier transform, and the solutions will in these cases be ordinary functions - although in some cases these functions do not have classical Fourier transforms.

In addition to consistency, we will note that there in some cases are analogous uniqueness results for the distributional solutions of differential equation as there are in the classical cases, in particular for linear first order ordinary differential equations \([6]\).

**Proposition 4.2** (Exercise 1, Chapter 5 in [5], Estimates for the Laplace and heat equation in the half space). For bounded solutions \(u(x,t)\) to the Laplace equation

\[ \Delta u(x,t) = 0 \]  

and to the heat equation

\[ \frac{\partial}{\partial t} u(x,t) = k\Delta_x u(x,t) \]

with \(k > 0\), both considered in the half-space \(x \in \mathbb{R}^n, t \geq 0\), with initial condition \(u(x,0) = f(x) \in \mathcal{S}\), we have

\[ \int_{\mathbb{R}^n} |u(x,t)|^2 \, dx \leq \int_{\mathbb{R}^n} |u(x,0)|^2 \, dx, \quad t > 0 \]  

Moreover, for \(u\) in either of (16) or (17) under these conditions, we have

\[ \lim_{t \to 0} \int_{\mathbb{R}^n} |u(x,t)|^2 \, dx = \int_{\mathbb{R}^n} |u(x,0)|^2 \, dx \quad \text{and} \quad \lim_{t \to \infty} \int_{\mathbb{R}^n} |u(x,t)|^2 \, dx = 0 \]

**Proof.** First the Laplace equation. Rewriting equation (16) slightly as

\[ \Delta u(x,t) = \left[ \frac{\partial^2}{\partial t^2} + \Delta_x \right] u(x,t) = 0 \]

and then applying the Fourier transform in the \(x\)-variables one obtains

\[ \mathcal{F}_x \frac{\partial^2 u}{\partial t^2} (\xi, t) - |\xi|^2 \mathcal{F}_x u(\xi, t) = 0 \]

which defines a second order ordinary differential equation in the \(t\)-variable with characteristic equation \(r^2 - |\xi|^2 = 0\) with solutions \(r_{1,2} = \pm |\xi|\). The classical
solution is thus

\[ \mathcal{F}_x u(\xi, t) = c_1(\xi)e^{t|\xi|} + c_2(\xi)e^{-t|\xi|} \]

where the constants in terms of \( t \), \( c_1(\xi) \) and \( c_2(\xi) \), in general may be functions of \( \xi \).

Now, because \( u(x, t) \) was assumed to be bounded, \( c_1(\xi) = 0 \ \forall \xi \in \mathbb{R}^n \) because \( e^{t|\xi|} \) is unbounded with respect to \( t > 0 \), and so it would make the Fourier transform in \( x \) of \( u(x, t) \) unbounded with respect to \( t \) and in turn \( u(x, t) \) would also be, contrary to the assumption. Thus the above equation reduces to

\[ \mathcal{F}_x u(\xi, t) = c_2(\xi)e^{-t|\xi|} \]

which together with the initial condition becomes \( \mathcal{F}_x u(\xi, 0) = \mathcal{F} f(\xi) = c_2(\xi)e^{-t|\xi|} \|_{t=0} = c_2(\xi) \). Thus it follows that

\[
\int_{\mathbb{R}^n} |u(x, t)|^2 \, dx \overset{(1)}{=} (2\pi)^{-n} \int_{\mathbb{R}^n} |c_2(\xi)e^{-t|\xi|}|^2 \, d\xi \\
\leq \overset{(2)}{(2\pi)^{-n} \int_{\mathbb{R}^n} |c_2(\xi)|^2 \, d\xi} \\
= (2\pi)^{-n} \int_{\mathbb{R}^n} |\mathcal{F}_x u(\xi, 0)|^2 \, d\xi \\
\overset{(1)}{=} \int_{\mathbb{R}^n} |u(\xi, 0)|^2 \, d\xi
\]

which proves the first statement for (16). (1) indicates application of Plancherel’s formula (Theorem 2.24), and in (2) we used the fact that \( e^{-t|\xi|} \leq 1 \) for \( t \geq 0 \). In the limits \( t \to 0 \) and \( t \to \infty \) we use the expression in the first equality above to obtain

\[
\lim_{t \to 0} \int_{\mathbb{R}^n} |u(x, t)|^2 \, dx = \lim_{t \to 0} (2\pi)^{-n} \int_{\mathbb{R}^n} |c_2(\xi)e^{-t|\xi|}|^2 \, d\xi \\
\overset{(\ast)}{=} (2\pi)^{-n} \int_{\mathbb{R}^n} |c_2(\xi)|^2 \, d\xi \\
= (2\pi)^{-n} \int_{\mathbb{R}^n} |\mathcal{F} f(\xi)|^2 \, d\xi \\
\overset{(1)}{=} \int_{\mathbb{R}^n} |f(x)|^2 \, dx \\
= \int_{\mathbb{R}^n} |u(x, 0)|^2 \, dx
\]
\[ \lim_{t \to \infty} \int_{\mathbb{R}^n} |u(x, t)|^2 \, dx = \lim_{t \to \infty} (2\pi)^{-n} \int_{\mathbb{R}^n} |c_2(\xi)e^{-t|\xi|^2}|^2 \, d\xi. \]

\[ \overset{(**) \text{Plancherel's formula (Theorem 2.24)}}{=} \int_{\mathbb{R}^n} |0|^2 \, d\xi = 0 \]

where (1) again indicates application of Plancherel’s formula (Theorem 2.24), and the limits may be interchanged with the integrals in (*) and (**) by noting that

\[ |c_2(\xi)e^{-t|\xi|^2}|^2 \leq |c_2(\xi)|^2 = |\mathcal{F}f(\xi)|^2 \in L^1 \]

due to the fact that \( f \in \mathcal{S} \) and then applying Lebesgue’s theorem of dominated convergence (Theorem 2.6). This concludes the proof for the Laplace equation.

Next, the heat equation. Applying the Fourier transform in the \( x \)-variable on equation (17) we get

\[ \frac{\partial}{\partial t} \mathcal{F}_x u(\xi, t) = -k |\xi|^2 \mathcal{F}_x u(\xi, t)) \]

Again, this is an ordinary differential equation in \( t \) whose classical solutions are given by

\[ \mathcal{F}_x u(\xi, t) = c(\xi)e^{-k|\xi|^2t} \]

for some function \( c(\xi) \). Applying the initial condition,

\[ \mathcal{F}_x u(\xi, 0) = \mathcal{F}f(\xi) = c(\xi) \]

Now we may immediately write

\[ \int_{\mathbb{R}^n} |u(x, t)|^2 \, dx \overset{(1)}{=} (2\pi)^{-n} \int_{\mathbb{R}^n} |c(\xi)e^{-k|\xi|^2t}|^2 \, d\xi \]

\[ \overset{(2)}{\leq} (2\pi)^{-n} \int_{\mathbb{R}^n} |c(\xi)|^2 \, d\xi \]

\[ = (2\pi)^{-n} \int_{\mathbb{R}^n} |\mathcal{F}_x u(\xi, t)|^2 \, dx \]

\[ \overset{(1)}{=} \int_{\mathbb{R}^n} |u(x, 0)|^2 \, dx \]

where once again (1) indicates application of Plancherel’s formula (Theorem 2.24) and (2) indicates use of the inequality \( e^{-k|\xi|^2t} \leq 1 \) for \( t > 0 \). For the limits, we may precisely as for the Laplace equation proceed by application of Theorem 2.6 in (*)
and (**) relying on the bound \( |c(\xi)e^{-k|\xi|^2t}|^2 \leq |c(\xi)|^2 = |\mathcal{F}f(\xi)|^2 \), which yields
\[
\lim_{t \to 0} \int_{\mathbb{R}^n} |u(x,t)|^2 \, dx = \lim_{t \to 0} (2\pi)^{-n} \int_{\mathbb{R}^n} \left| c(\xi)e^{-k|\xi|^2t} \right|^2 \, d\xi = (2\pi)^{-n} \int_{\mathbb{R}^n} |c(\xi)|^2 \, d\xi
\]
\[
= (2\pi)^{-n} \int_{\mathbb{R}^n} |\mathcal{F}_x u(\xi,0)|^2 \, d\xi
\]
\[
= \int_{\mathbb{R}^n} |u(x,0)|^2 \, dx
\]
and
\[
\lim_{t \to \infty} \int_{\mathbb{R}^n} |u(x,t)|^2 \, dx = \lim_{t \to \infty} (2\pi)^{-n} \int_{\mathbb{R}^n} \left| c(\xi)e^{-k|\xi|^2t} \right|^2 \, d\xi = \int_{\mathbb{R}^n} |0|^2 \, d\xi = 0
\]
and the proof for the heat equation is complete. \( \square \)

**Proposition 4.3** (Exercise 2, Chapter 5 in [5], More Estimates for the Laplace and Heat equation in the half space). For solutions \( u(x,t) \) to the equations (16) and (17) as previously, we have the following estimate
\[
|u(x,t)| \leq \sup_{y \in \mathbb{R}^n} |u(y,0)|
\]

**Proof.** For (16), we have from before that
\[
\mathcal{F}_x u(\xi,t) = \mathcal{F}f(\xi)e^{-t|\xi|}
\]
Applying the inverse Fourier transform and using Proposition 3.25 we obtain
\[
u(x,t) = \mathcal{F}^{-1}_x \left( \mathcal{F}_x f(\xi) \cdot e^{-t|\xi|} \right)(x,t) = f \ast \mathcal{F}^{-1}_x e^{-t|\xi|}(x,t)
\]
This gives, by the triangle inequality
\[
|u(x,t)| \leq \int_{\mathbb{R}^n} |f(x-y)| \left| \mathcal{F}^{-1}_x e^{-t|\xi|} (y,t) \right| \, dy
\]
(19)
We now utilize the fact that (see [5])
\[
\mathcal{F}^{-1}_x e^{-t|\xi|}(x,t) = \pi^{-\frac{n-1}{2}} \Gamma \left( \frac{n+1}{2} \right) \frac{t}{(t^2 + |x|^2)^{\frac{n+1}{2}}} \geq 0
\]
for $t > 0, x \in \mathbb{R}^n$, and for each $t > 0$ this is in $L^1(\mathbb{R}^n)$ w.r.t. $x$ since it is bounded on compact sets and decays as $|x|^{-n-1}$ when $|x|$ grows large. From this we have that for $t > 0$,

$$\int_{\mathbb{R}^n} \mathcal{F}_x^{-1} e^{-t|\xi|} (x, t) \, dx = \mathcal{F}_x (\mathcal{F}_x^{-1} e^{-t|\xi|}) (0, t) = e^{-0t} = 1$$

and since $f \in \mathcal{S}$, $f$ is bounded. Hence from (19) we have

$$|u(x, t)| \leq \int_{\mathbb{R}^n} |f(x - y)| \left| \mathcal{F}_x^{-1} e^{-t|\xi|} \right| \, dy$$

$$\leq \int_{\mathbb{R}^n} \sup_{y \in \mathbb{R}^n} |f(x - y)| \left| \mathcal{F}_x^{-1} e^{-t|\xi|} \right| \, dy$$

$$= \sup_{y \in \mathbb{R}^n} |f(y)| \int_{\mathbb{R}^n} \mathcal{F}_x^{-1} e^{-t|\xi|} \, dy$$

$$= \sup_{y \in \mathbb{R}^n} |u(y, 0)|$$

since $f(y) = u(y, 0)$ and the proof for (16) is done.

Now turning to (17) we instead have

$$\mathcal{F}_x u(\xi, t) = \mathcal{F} f(\xi) e^{-k|\xi|^2 t}$$

which by an analogous procedure, applying $\mathcal{F}^{-1}$ and using Proposition 3.25 becomes

$$u(x, t) = \mathcal{F}_x^{-1} \left( \mathcal{F} f(\xi) e^{-k|\xi|^2 t} \right) (x, t) = f * \mathcal{F}^{-1} (e^{-k|\xi|^2 t}) (x, t)$$

we note that $e^{-k|\xi|^2 t} = e^{-\frac{k^2}{4t} x^2}$ with $a = \frac{4}{kt}$ so that by Example 2.7

$$\mathcal{F}_x^{-1} (e^{-k|\xi|^2 t}) (x, t) = \left( \sqrt{\frac{kt}{4\pi}} \right)^n e^{-\frac{4x^2}{kt}} \geq 0$$

Also as in the previous case

$$\int_{\mathbb{R}^n} \mathcal{F}_x^{-1} (e^{-k|\xi|^2 t}) (x, t) \, dx = \mathcal{F}_x \mathcal{F}_x^{-1} (e^{-k|\xi|^2 t}) (0, t) = e^{-k|0|^2 t} = 1$$

We thus as in the previous case have

$$|u(x, t)| \leq (1) \int_{\mathbb{R}^n} |f(x - y)| \left| \mathcal{F}^{-1} e^{-k|\xi|^2 t} (y, t) \right| \, dy$$

$$\leq \sup_{y \in \mathbb{R}^n} |f(y)| \int_{\mathbb{R}^n} \mathcal{F}_x^{-1} e^{-k|\xi|^2 t} (y, t) \, dy$$

$$= \sup_{y \in \mathbb{R}^n} |u(y, 0)|$$
using the facts above, the triangle inequality for integrals (1) and the boundedness of $f$ (2).

\[\square\]

**Proposition 4.4** (Exercises 10 and 11, Chapter 5 in [5], Klein-Gordon equation, Fourier transform of the solution and conservation of energy). The solution $u(x, t)$, $(x, t) \in \mathbb{R}^n \times \mathbb{R}$ to the Klein-Gordon equation

\[\frac{\partial^2}{\partial t^2} u = \Delta_x u - m^2 u, \quad m > 0\quad (20)\]

with initial conditions

\[u(x, 0) = f(x) \quad \text{and} \quad \frac{\partial u}{\partial t}(x, 0) = g(x)\quad (21)\]

for some functions $f, g \in \mathcal{S}(\mathbb{R}^n)$ satisfies

\[\mathcal{F}_x u(\xi, t) = \mathcal{F}_x g(\xi) \sin(t \sqrt{|\xi|^2 + m^2}) + \mathcal{F}_x f(\xi) \cos(t \sqrt{|\xi|^2 + m^2})\]

Moreover, the energy $E(t)$ of the solution $u$ defined as

\[E(t) := \frac{1}{2} \int_{\mathbb{R}^n} \left( m^2 |u(x, t)|^2 + \left| \frac{\partial}{\partial t} u(x, t) \right|^2 + \sum_{i=1}^{n} \left| \frac{\partial}{\partial x_i} u(x, t) \right|^2 \right) \, dx \quad (22)\]

is independent of time.

**Proof.** Applying $\mathcal{F}_x$ to \[(20)\] we obtain

\[\frac{\partial^2}{\partial t^2} \mathcal{F}_x u(\xi, t) = - (|\xi|^2 + m^2) \mathcal{F}_x u(\xi, t)\]

Viewed classically, this is an ordinary differential equation of order 2, with characteristic equation $r^2 + (|\xi|^2 + m^2) = 0$ which have the roots $r = \pm \sqrt{- |\xi|^2 - m^2} = \pm i \sqrt{|\xi|^2 + m^2}$, and thus have classical solutions of the form

\[\mathcal{F}_x u(\xi, t) = c_1(\xi) e^{it \sqrt{|\xi|^2 + m^2}} + c_2(\xi) e^{-it \sqrt{|\xi|^2 + m^2}}\]

Any such function define a tempered distribution as long as $c_1, c_2$ grows at most polynomially, since the exponentials are bounded. Applying $\mathcal{F}$ to the initial conditions and equating it with $\mathcal{F}_x u$ as a distribution on the form above, yields

\[\mathcal{F}_x u(\xi, 0) = c_1(\xi) + c_2(\xi) = \mathcal{F}_x f(\xi)\]

\[\frac{\partial}{\partial t} \mathcal{F}_x u(\xi, 0) = i \sqrt{|\xi|^2 + m^2} (c_1(\xi) - c_2(\xi)) = \mathcal{F}_x g(\xi)\]
Solving this system of equations for $c_1(\xi), c_2(\xi)$ gives

$$c_1(\xi) = \frac{1}{2iD(\xi)} \mathcal{F}_x g(\xi) + \frac{1}{2} \mathcal{F}_x f(\xi)$$
$$c_2(\xi) = \frac{1}{2} \mathcal{F}_x f(\xi) - \frac{1}{2iD(\xi)} \mathcal{F}_x g(\xi)$$

where we have put $D(\xi) = \sqrt{\left|\xi\right|^2 + m^2}$. Putting this into the equation for $\mathcal{F}_x u(\xi, t)$,

$$\mathcal{F}_x u(\xi, t) = \frac{1}{D(\xi)} \mathcal{F}_x g(\xi) \left( \frac{e^{itD(\xi)} - e^{-itD(\xi)}}{2i} \right) + \mathcal{F}_x f(\xi) \left( \frac{e^{itD(\xi)} + e^{-itD(\xi)}}{2} \right)$$
$$= \frac{\mathcal{F}_x g(\xi)}{D(\xi)} \sin (tD(\xi)) + \mathcal{F}_x f(\xi) \cos ((tD(\xi)))$$
$$= \frac{\mathcal{F}_x g(\xi)}{\sqrt{\left|\xi\right|^2 + m^2}} \sin \left( t\sqrt{\left|\xi\right|^2 + m^2} \right) + \mathcal{F}_x f(\xi) \cos \left( t\sqrt{\left|\xi\right|^2 + m^2} \right)$$

which was the first claim. Now this means

$$\left|\mathcal{F}_x u(\xi, t)\right|^2 \, dx = \left( \frac{\mathcal{F}_x g(\xi)}{D(\xi)} \sin (tD(\xi)) + \mathcal{F}_x f(\xi) \cos (tD(\xi)) \right)^2$$
$$= \frac{\left|\mathcal{F}_x g(\xi)\right|^2}{(D(\xi))^2} \sin^2 (tD(\xi))$$
$$+ \frac{2 \left|\mathcal{F}_x g(\xi)\right| \left|\mathcal{F}_x f(\xi)\right|}{D(\xi)} \sin (tD(\xi)) \cos (tD(\xi))$$
$$+ \left|\mathcal{F}_x f(\xi)\right|^2 \cos^2 (tD(\xi))$$

(23)

and from the expression for $\mathcal{F}_x u$,

$$\frac{\partial}{\partial t} \mathcal{F}_x u(\xi, t) = \mathcal{F}_x g(\xi) \cos (tD(\xi)) - \mathcal{F}_x f(\xi) D(\xi) \sin (tD(\xi))$$

so that

$$\left|\frac{\partial}{\partial t} \mathcal{F}_x u(\xi, t)\right|^2 = \left|\mathcal{F}_x g(\xi)\right|^2 \cos^2 (tD(\xi))$$
$$- 2 \left|\mathcal{F}_x g(\xi)\right| \left|\mathcal{F}_x f(\xi)\right| D(\xi) \sin (tD(\xi)) \cos (tD(\xi))$$
$$+ \left|\mathcal{F}_x f(\xi)\right|^2 (D(\xi))^2 \sin^2 (tD(\xi))$$

(24)
Note that the mixed terms in equations (23) and (24) differ with precisely a factor of \(- (D(\xi))^2 = -(|\xi|^2 + m^2)\), and therefore

\[
(D(\xi))^2 |F_x u(\xi, t)|^2 + \left| \frac{\partial}{\partial t} F_x u(\xi, t) \right|^2 = |F_x g(\xi)|^2 \sin^2 (tD(\xi)) + |F_x f(\xi)|^2 (D(\xi))^2 \cos^2 (tD(\xi)) + |F_x g(\xi)|^2 \cos^2 (tD(\xi)) + |F_x f(\xi)|^2 (D(\xi))^2 \sin^2 (tD(\xi)) = |F_x g(\xi)|^2 + |F_x f(\xi)|^2 (D(\xi))^2
\]

(25)

by grouping the terms and using \(\sin^2 x + \cos^2 x = 1\). Also, by \(F_x \left( \frac{\partial}{\partial x_i} u \right)(\xi, t) = -i\xi F_x u(\xi, t)\) from Proposition \(\ref{eq:prop3.25}\), we have

\[
\sum_{i=1}^{n} \left| F_x \left( \frac{\partial}{\partial x_i} u \right)(\xi, t) \right|^2 = |\xi|^2 |F_x u(\xi, t)|^2
\]

(26)

Now, if we apply Plancherel’s formula to (22), we may write

\[
E(t) = \frac{1}{2(2\pi)^n} \int_{\mathbb{R}^n} \left( m^2 |F_x u(\xi, t)|^2 + \left| \frac{\partial}{\partial t} F_x u(\xi, t) \right|^2 + \sum_{i=1}^{n} \left| F_x \left( \frac{\partial}{\partial x_i} u \right)(\xi, t) \right|^2 \right) d\xi
\]

(25)

\[
= \frac{1}{2(2\pi)^n} \int_{\mathbb{R}^n} \left( |\xi|^2 + m^2 \right) |F_x u(\xi, t)|^2 + \left| \frac{\partial}{\partial t} F_x u(\xi, t) \right|^2 \right) d\xi
\]

(26)

\[
= \frac{1}{2(2\pi)^n} \int_{\mathbb{R}^n} |F_x g(\xi)|^2 + |F_x f(\xi)|^2 (|\xi|^2 + m^2) \ d\xi
\]

where the last expression is clearly independent of \(t\), and the second claim is proven.

\[\square\]

As a theoretical chemist, the author finds the following, final example especially interesting and enjoyable, as it concerns the Schrödinger equation - which essentially is the fundamental equation in quantum mechanics that underpins the quantum theory of chemistry. Albeit the free Schrödinger equation is considered here, while in chemistry the vast majority of interesting systems such as molecules are bound states which includes a non-trivial potential term in the equation, a connection still exists of course.

**Proposition 4.5** (Exercise 19, Chapter 5 in [5], Limiting behavior of \(L^1\) solutions to the free Schrödinger equation). For the solution \(u(x, t)\) of the free Schrödinger
equation (27)

(27) \[ \frac{\partial}{\partial t} u(x, t) = i k \nabla_x^2 u(x, t) \]

for \( x \in \mathbb{R}^3, \ t \geq 0, \) and with initial condition \( u(x, t) = \varphi(x) \) where

(28) \[ \int_{\mathbb{R}^3} |\varphi(x)| \, dx < \infty \]

it holds that

(29) \[ |u(x, t)| \leq M t^{-\frac{3}{2}} \]

for some constant \( M \in \mathbb{R}. \) As a consequence,

\[ \int_B |u(x, t)|^2 \, dx \to 0 \]

as \( t \to \infty \) for any bounded region \( B \subset \mathbb{R}^3. \)

**Proof.** Applying the Fourier transform in the \( x \)-variable to (27) yields, after application of Proposition 3.25,

(30) \[ \frac{\partial}{\partial t} \mathcal{F}_x u(\xi, t) = -i k |\xi|^2 \mathcal{F}_x u(\xi, t) \]

which again is an ordinary differential equation in the \( t \)-variable with solution

(31) \[ \mathcal{F}_x u(\xi, t) = c(\xi) e^{i k |\xi|^2 t} \]

and by taking the Fourier transform of the initial condition we obtain

\[ c(\xi) = \mathcal{F}_x u(\xi, 0) = \mathcal{F}_x \varphi(\xi) \]

Inserting this into equation (31), and then taking the inverse Fourier transform, followed by application of Proposition 3.25 leads to

(32) \[ u(x, t) = \mathcal{F}_x^{-1} \left( \mathcal{F}_x \varphi(\xi) e^{i k |\xi|^2 t} \right)(x, t) = \varphi \ast \mathcal{F}_x^{-1} \left( e^{i k |\xi|^2 t} \right)(x, t) \]

We now make use of the following fact (see [5])

\[ \mathcal{F}_x e^{i k |x|^2 t}(\xi, t) = \left( \frac{\pi}{kt} \right)^{\frac{3}{2}} e^{i\pi\frac{3}{4}} e^{-\frac{|\xi|^2}{4t}} \]

where we have assumed \( k > 0, \ t > 0. \) This implies from Proposition 3.25 that

(33) \[ \mathcal{F}_x^{-1} e^{i k |\xi|^2 t}(\xi, t) = (2\pi)^{-\frac{3}{2}} \left( \frac{\pi}{kt} \right)^{\frac{3}{4}} e^{i\pi\frac{3}{4}} e^{-\frac{|\xi|^2}{4t}} \]
Now we easily obtain by combining equations (32) and (33) that
\[
|u(x,t)| = \left| \int_{\mathbb{R}^3} \varphi(x-y) (2\pi)^{-3} \left( \frac{\pi}{kt} \right)^{\frac{3}{2}} e^{i\pi \frac{3}{4} \frac{y^2}{kt}} \right|
\leq \left| (2\pi)^{-3} \left( \frac{\pi}{kt} \right)^{\frac{3}{2}} \int_{\mathbb{R}^3} |\varphi(x-y)| \, dy \right|
\leq M |t|^{-\frac{3}{2}}
\]
where
\[
M = 2^{-3} \pi^{-\frac{3}{2}} \| \varphi \|_{L^1} < \infty
\]
is a constant independent of \(x\) and \(t\). Thus the first claim is proven. The consequence is easy to show, since if \(B \subset \mathbb{R}^3\) is bounded, then we can evidently estimate
\[
\int_B |u(x,t)|^2 \, dx \leq M^2 |t|^{-3} \int_B 1 \, dx
\]
where \(|B| < \infty\) is the volume (measure) of \(B\), and the last expression evidently approaches 0 as \(t \to \infty\).

The \(L^2\)-norm over a set \(B\) of the solution to the Schrödinger equation is in quantum mechanics usually interpreted as the probability of finding the particle in the region \(B\) (up to a constant if the \(L^2(\mathbb{R}^3)\)-norm of the solution is not normalized). Hence the above example shows that for a particle described by an initially integrable solution to the free Schrödinger equation, the probability of finding the particle in any bounded region vanishes as time passes.
REFERENCES


