university of groningen

faculty of science and engineering

mathematics and applied mathematics

Distribution Theory

Bachelor's Project Mathematics

July 2018

Student: N.B-S. Boer

First supervisor: Dr. A.E. Sterk

Second assessor: Dr.ir. R. Luppes

Abstract

Distribution theory is a very broad field in mathematics, that can be used to solve a wide range of applications, mainly those involving differential equations. In this report, distributions will be defined and a broad theoretical basis of distribution theory will be laid. This is necessary in order to understand how distributions can be used. The goal of this thesis is to explore its applications, and the last two chapters will be focusing on this.

Contents

1. Introduction

When working on applied mathematical problems, especially ones related to differential equations, one may often encounter a situation in which a discontinuous function appears in its solution. In this case, the answer to a differential equation made up of completely continuous functions and common differential operators, suddenly yields a discontinuous solution. At times, this can make one wonder; why does this occur, can this even be a valid solution, and how to even define differentiation on a discontinuous function? Usually, this problem is worked around by using some step function, or a Dirac delta, and saying that if the differential equation is integrated, somehow these functions satisfy it. However, this is not a very satisfying answer, and this is the place where distribution theory comes in.

On the other hand, comparable situations may occur in measure theory, something described elaborately in [1]. Why exactly does Fubini's theorem only work for functions that satisfy certain conditions? It turns out that this restriction is not necessary at all, as one can greatly extend a lot of theorems from this field by using a broader definition of what a function actually means, and that too is the place where distribution theory comes in. Then there is a whole field of mathematical problems, that is often tackled by applying a Fourier transformation. It will turn out that this notion too, can be greatly generalized, to act on all kinds of maps that some may not even consider functions. This is especially a field where distribution theory comes in, and will in fact be one of the main goals of this report.

In order to fully explain why distribution theory is involved in all of these scenarios, one will first have to lay a solid foundation of the theory behind distributions, and explore theoretical notions in order to grasp the great multitude of applications. This is why the lion's share of this report will be spent on the theory behind distributions, and the specific applications in will only be treated in the last few chapters. However, let us also in the theoretical parts focus on how every notion will be useful in an applied setting, in order to remember the goal ahead; applications of distribution theory.

Though, exactly how can distribution theory fill in the gaps in all these courses, and in those applications? First of all, in distribution theory one works with so-called test functions. These are not the broadly defined functions announced earlier, on the contrary, these are infinitely differentiable functions with a compact support. Exactly because these functions are simple to work with, for example integration of such function is always well-defined, this type of functions is used to define distributions. These distributions are sometimes called generalized functions, and for good reason. While every function is also a distribution, for the usual applications one considers more irregular distributions. In fact, it will turn out that distributions need to satisfy almost no properties.

Then one may ask oneself, what use are distributions when they need to satisfy that few properties? Well it turns out, that there is still a lot to extract from the information how a distribution would act in a certain situation. This, in fact, will be a recurring theme throughout this report. Let us not skip ahead of ourselves however, and carefully define and explore test functions, starting next chapter.

2. Test Functions

2.1. Appearance of Test Functions

In order to fully understand the necessity and use of a so-called test function, let us consider the following simple ordinary differential equation:

$$x \cdot u'(x) = 0, \text{ where } x \in \mathbb{R}.$$

Classically, in order for this equation to hold at all points $x \in \mathbb{R}$, $u(x)$ has to be constant, except at $x = 0$, for which no solution of $u(x)$ can be found. However, one can retrieve more information about this function, and even extend the solution to be defined at $x = 0$. Consider the weak form of this differential equation, which is that the inner product $\langle x \cdot u'(x), \varphi(x)\rangle = 0$ for a continuously differentiable real function $\varphi(x)$ that is zero outside some bounded interval of $\mathbb{R}$. Recall that the inner product of functions on $\mathbb{R}$ is given by

$$\langle f, g\rangle \equiv \int_{-\infty}^{\infty} f(x)\ g(x)\ dx,$$

given that this integral exists. Hence the weak form of the differential equation can be rewritten as:

$$\begin{aligned}
\langle x \cdot u'(x), \varphi(x)\rangle &= \int_{-\infty}^{\infty} x\ u'(x)\ \varphi(x)\ dx \\
&= -\int_{-\infty}^{\infty} x\ u(x)\ \varphi'(x)\ dx + \int_{-\infty}^{\infty} x \cdot (u(x) \cdot \varphi(x))'\ dx \\
&= -\int_{-\infty}^{\infty} x\ u(x)\ \varphi'(x)\ dx - \int_{-\infty}^{\infty} u(x)\ \varphi(x)\ dx \\
&= -\langle u(x), \varphi(x) + x \cdot \varphi'(x)\rangle.
\end{aligned}$$

Consider the Heaviside function $H(x) = \begin{cases} 1 \text{ for } x \geq 0, \\ 0 \text{ for } x < 0. \end{cases}$ Then for $x \geq 0$:

$$\begin{aligned}
\langle x \cdot H'(x), \varphi(x)\rangle &= -\langle H(x), \varphi(x) + x \cdot \varphi'(x)\rangle \\
&= -\langle 1, \varphi(x) + x \cdot \varphi'(x)\rangle \\
&= -\int_{0}^{\infty} \varphi(x) + x\ \varphi'(x)\ dx \\
&= -\int_{0}^{\infty} \varphi(x)\ dx + \int_{0}^{\infty} \varphi(x)\ dx = 0,
\end{aligned}$$

using partial integration on the second and third line. Hence $H(x)$ clearly satisfies the differential equation, as $H(x) = 0$ for $x < 0$, which trivially satisfies it as well. Therefore, the general solution of the differential equation is given by $u(x) = c_1 \cdot H(x) + c_2$, for some constants $c_1, c_2 \in \mathbb{R}$. Not only does this give us a bit more information about the solution itself, but it also defines a solution in the case that $x = 0$.

Note that while $\varphi(x)$ was essential to deriving the solution, the function itself does not actually appear in it. This is an example of a test function; functions that need not be explicitly defined, but can be used to show a certain effect or property of another function or operation acting upon it. A more restrictive and mathematically sound definition will be given next section.

For now, let us use this setting to both take a look at the appearance of distributions, and see how test functions are involved in such example. Consider an often used model in quantum mechanics; a potential of the form $V = \frac{1}{r}$ embedded in $\mathbb{R}^3$, with r the distance to the origin. Outside of the origin, $\Delta V = 0$, but ΔV cannot conventionally be defined at the origin. However, one can express the value at the origin by setting $\Delta V = -4\pi \cdot \delta$, where δ is the so-called Dirac delta. Informally, the Dirac delta is usually defined as

$$\delta(x) = \begin{cases} 0 \text{ for } x \neq 0, \\ \infty \text{ for } x = 0. \end{cases}$$

With the use of test functions however, it can be defined mathematically by $\langle \delta(x), \varphi(x) \rangle \equiv \varphi(0)$, for any real valued test function $\varphi(x)$. The last phrase however, 'for any test function $\varphi(x)$', raises the question which will form the first topic of interest; which functions count as test functions, and how are test functions mathematically defined?

2.2. Test Functions

In order to define a test function, let us first introduce two others notions.

Definition 2.1. Given $\varphi : \mathbb{R}^n \to \mathbb{C}$, the support of φ is defined as $\text{supp}(\varphi) \equiv \overline{\{x \in \mathbb{R}^n \mid \varphi(x) \neq 0\}}$.

Note that since the support is the closure of the set of nonzero points of φ, it is a closed set in $\mathbb{R}^n$. Recall that a subset of $\mathbb{R}^n$ is compact if and only if it is closed and bounded. Hence whenever the set of nonzero points of φ is bounded, its support is bounded and thus compact.

Definition 2.2. Let $k = (k_1, \ldots, k_n) \in \mathbb{N}^n$. Then the partial differential operator is defined as

$$D^k \equiv \left(\frac{\partial}{\partial x_1}\right)^{k_1} \cdots \left(\frac{\partial}{\partial x_n}\right)^{k_n} = \frac{\partial^{|k|}}{\partial x_1^{k_1} \cdots \partial x_n^{k_n}},$$

where $|k| \equiv k_1 + \cdots + k_n$ is the order of D^k and $D^0 \equiv id$.

Now all the tools are in place to define both test functions, and the space of test functions.

Definition 2.3. A function $\varphi : \mathbb{R}^n \to \mathbb{C}$ is said to be a C^m function if $D^k\varphi$ exists and is continuous for all $k \in \mathbb{N}^n$ such that $|k| \leq m$. The space of all C^m functions is denoted $\mathcal{E}^m(\mathbb{R}^n)$. Furthermore, φ is a C^∞ function if it is a C^m function for all $m \in \mathbb{N}$, i.e. whenever $D^k\varphi$ exists and is continuous for all $k \in \mathbb{N}^n$. The space of all C^∞ functions is denoted $\mathcal{E}(\mathbb{R}^n)$.

Now consider $\mathcal{D}(\mathbb{R}^n) \equiv \{\varphi \in \mathcal{E}(\mathbb{R}^n) \mid \text{supp}(\varphi) \subset \mathbb{R}^n \text{ is compact}\}$. All elements $\varphi \in \mathcal{D}(\mathbb{R}^n)$, i.e. C^∞ functions with compact support, are called test functions.

In notation, $\mathcal{D}(\mathbb{R}^n)$ and $\mathcal{E}(\mathbb{R}^n)$ are often denoted $\mathcal{D}$ and $\mathcal{E}$ respectively. Both sets can be restricted by defining $\mathcal{E}(\Omega) \equiv \{\varphi \in \mathcal{E}(\mathbb{R}^n) \mid \text{dom}(\varphi) \subset \Omega\}$, for some open set $\Omega \subset \mathbb{R}^n$, and $\mathcal{D}(\Omega) \equiv \{\varphi \in \mathcal{D}(\mathbb{R}^n) \mid \text{dom}(\varphi) \subset \Omega\} = \{\varphi \in \mathcal{E}(\Omega) \mid \text{supp}(\varphi) \subset \Omega \text{ is compact}\}$.

To avoid confusion, let us abbreviate $\mathcal{D}(\Omega)$ and $\mathcal{E}(\Omega)$ to $\mathcal{D}$ and $\mathcal{E}$ respectively only if $\Omega = \mathbb{R}^n$.

It is now useful to investigate some properties of $\mathcal{D}$, in order to get a better sense of this space.

Property 2.1. $\mathcal{D}$ is a linear space.

Proof. Let $\varphi, \psi \in \mathcal{D}$, then both φ and ψ are C^∞ functions with compact support. Hence:

1. Since $\text{supp}(\varphi+\psi) \subset \text{supp}(\varphi) \cup \text{supp}(\psi)$, and the latter is a compact set, $\text{supp}(\varphi+\psi)$ must be compact too. Moreover, since φ and ψ are C^∞ functions, their sum is a C^∞ function as well. Then by Definition 2.3, $\varphi + \psi \in \mathcal{D}$.

2. Let $\lambda \in \mathbb{C}$, then $\text{supp}(\lambda \cdot \varphi) = \text{supp}(\varphi)$. Also since $\lambda \cdot \varphi$ is a C^∞ function, by Definition 2.3 $\lambda \cdot \varphi \in \mathcal{D}$.

Therefore, $\mathcal{D}$ is a linear space. □

Property 2.2. Let $\varphi \in \mathcal{D}$, then:

1. $D^k \varphi \in \mathcal{D}\ \forall k \in \mathbb{N}^n$.

2. $f \cdot \varphi \in \mathcal{D}\ \forall f \in \mathcal{E}$.

Proof. Since $\varphi \in \mathcal{D}$, it is a C^∞ function with compact support. Now let us prove the properties:

1. This follows immediately from the observation that for all $k \in \mathbb{N}^n, \text{supp}(D^k\varphi) \subset \text{supp}(\varphi)$. Since $\text{supp}(\varphi)$ is compact, $\text{supp}(D^k\varphi)$ must be as well, and thus $D^k\varphi \in \mathcal{D}$ for all $k \in \mathbb{N}^n$.

2. By Definition 2.3 all $f \in \mathcal{E}$ are C^∞ functions and thus its multiplication by φ results in a C^∞ function. Moreover, $\text{supp}(f \cdot \varphi) \subset \text{supp}(f) \cap \text{supp}(\varphi) \subset \text{supp}(\varphi)$, therefore $\text{supp}(f \cdot \varphi)$ is compact as well and thus $f \cdot \varphi \in \mathcal{D}$ for all $f \in \mathcal{E}$. □

These properties will prove to be very useful when test functions will be used in the next chapters. However one more notion is to be introduced in order to fully lay the groundwork for using test functions. Property 2.2 in particular will allow us to follow up the next definition with a useful corollary.

Definition 2.4. A sequence $(\varphi_j) \in \mathcal{D}$ for $j \in \mathbb{N}$ is said to converge to $\varphi \in \mathcal{D}$ if there exists a compact set $K \subset \mathbb{R}^n$ such that:

1. $\text{supp}(\varphi_j) \subset K\ \forall j \in \mathbb{N}$.

2. $D^k\varphi_j \xrightarrow{\mathbb{C}} D^k\varphi$ uniformly, $\forall k \in \mathbb{N}^n$.

This is usually denoted $\varphi_j \xrightarrow{\mathcal{D}} \varphi$.

Note that here, and in all future cases, the arrow will indicate convergence as j goes to ∞. The space in which it converges will always be indicated above the arrow.

Corollary 2.5. Assume $\varphi_j \xrightarrow{\mathcal{D}} \varphi$, then:

1. $D^k\varphi_j \xrightarrow{\mathcal{D}} D^k\varphi\ \forall k \in \mathbb{N}^n$.

2. $f \cdot \varphi_j \xrightarrow{\mathcal{D}} f \cdot \varphi\ \forall f \in \mathcal{E}$.

Proof. Since $\varphi_j \xrightarrow{\mathcal{D}} \varphi$, by Definition 2.4 there exists a compact set $K \in \mathbb{R}^n$ such that $\text{supp}(\varphi_j) \subset K$ for all $j \in \mathbb{N}$. Now let us prove the claims:

1. Fix $k \in \mathbb{N}^n$, then for all $j, l \in \mathbb{N}$ the following holds: $\text{supp}(D^k\varphi_j) \subset \text{supp}(\varphi_j) \subset K$ as seen in the proof of Property 2.2, and from Definition 2.2 it follows that $D^l D^k = D^{l+k}$. Furthermore $D^l(D^k\varphi_j) = D^{l+k}\varphi_j$ converges uniformly to $D^{l+k}\varphi = D^l(D^k\varphi)$, and $D^{l+k}\varphi \in \mathbb{D}$ by Property 2.2. Since $k \in \mathbb{N}$ was chosen arbitrarily, this holds for all $k \in \mathbb{N}$, and thus by Definition 2.4, $D^k\varphi_j \xrightarrow{\mathcal{D}} D^k\varphi$ for all $k \in \mathbb{N}^n$.

2. Fix $f \in \mathcal{E}$, then for all $j \in \mathbb{N}$ and $k \in \mathbb{N}^n$ the following holds: $\text{supp}(f \cdot \varphi_j) \subset \text{supp}(\varphi_j) \subset K$ as seen in the proof of Property 2.2, and by Leibniz' formula

$$D^k(f \cdot \varphi_j) = \sum_{l=0}^{k} \binom{k}{l} \cdot D^l f \cdot D^{k-l}\varphi_j$$

   where $k! \equiv k_1! \cdots k_n!$. Hence $D^k(f \cdot \varphi_j)$ converges uniformly to $D^k(f \cdot \varphi)$. Since $f \in \mathcal{E}$ was chosen arbitrarily, this holds for all $f \in \mathcal{E}$, and thus by Definition 2.4, $f \cdot \varphi_j \xrightarrow{\mathcal{D}} f \cdot \varphi$ for all $f \in \mathcal{E}$. □

As a final note, and in preparation for the next chapter, the space $\mathcal{D}$ is sequentially complete with respect to the L^∞-norm. Although this will not be mentioned explicitly in later chapters, it is necessary in order to guarantee certain technicalities, especially with respect to convergence. The start of the next chapter will clarify immediately why this is the case. However, since the proof of this statement itself requires a lot of topological background that would have to be introduced, it will be left for the interested reader to investigate further in [2] (Proposition 1.8) and [3] (Theorem 1.22).

3. Distributions

3.1. Defining Distributions and Basic Properties

Now that test functions have been defined, and the properties of the test function space $\mathcal{D}$ have been explored, it is time to define distributions and most importantly explore their properties. After the definition, there will be a number of important examples. Nonetheless it may be helpful to recall while reading the definition, that in section 2.1 a distribution was already introduced; the Dirac delta. The properties and characteristics of the Dirac delta that one may be aware of, could help to understand why a distribution is defined in the following way.

Definition 3.1. A distribution is a map $T : \mathcal{D} \to \mathbb{C}$ such that for all $\varphi, \psi \in \mathcal{D}$ and $\lambda \in \mathbb{C}$:

1. $T(\varphi + \psi) = T(\varphi) + T(\psi)$.
2. $T(\lambda \cdot \varphi) = \lambda \cdot T(\varphi)$.
3. If $\varphi_j \xrightarrow{\mathcal{D}} \varphi$, then $T(\varphi_j) \xrightarrow{\mathbb{C}} T(\varphi)$.

This definition basically states that a distribution must be both linear and continuous with respect to test functions, and that the set of distributions is the dual space of $\mathcal{D}$, denoted $\mathcal{D}'$. The continuity is the condition for which it might help to think of a specific example, like the Dirac delta, in order to see why this is required. Later this chapter the reason for, and the consequences of, this continuity will be explored.

For now, note that the linearity is the reason that the distribution of a test function, for instance $T(\varphi)$, is usually denoted $\langle T, \varphi \rangle$ instead. However this notation suggests bi-linearity, which will therefore be shown immediately.

Property 3.1. $\mathcal{D}'$ is a linear space, where for $T_1, T_2 \in \mathcal{D}', \varphi \in \mathcal{D}$, and $\lambda \in \mathbb{C}$:

1. $\langle T_1 + T_2, \varphi \rangle = \langle T_1, \varphi \rangle + \langle T_2, \varphi \rangle$.
2. $\langle \lambda \cdot T_1, \varphi \rangle = \lambda \cdot \langle T_1, \varphi \rangle$.

Hence bi-linearity is now guaranteed, however it is crucial not to confuse this with the inner product, as $\langle T, \lambda \cdot \varphi \rangle = \lambda \cdot \langle T, \varphi \rangle$, and not multiplied by the complex conjugate as in the inner product. As will be shown shortly though, it turns out that there is a class of distributions for which these two notions *are* equal. However, let us not run ahead of things, and first consider the example seen before.

Example 3.2. The Dirac delta δ, defined as $\langle \delta, \varphi \rangle \equiv \varphi(0)$ for $\varphi \in \mathcal{D}$, is a distribution. Whenever its so-called delta spike is not at the origin, but at $x = a$, the corresponding Dirac delta is denoted $\delta_{(a)}$ and defined as $\langle \delta_{(a)}, \varphi \rangle \equiv \varphi(a)$ for $\varphi \in \mathcal{D}$. Note that throughout this report, $\delta_{(a)}$ will be used whenever a general Dirac delta is applicable, and δ only if $a = 0$ is necessary.

Now, the notion that a distribution can be written as an inner product of real functions, is generalized in the following example.

Example 3.3. Let $f : \mathbb{R}^n \to \mathbb{C}$ be locally integrable, then

$$\langle T_f, \varphi \rangle \equiv \int_{\mathbb{R}^n} f(x)\ \varphi(x)\ dx \text{ for } \varphi \in \mathcal{D}$$

is a distribution. Actually, any distribution $T \in \mathcal{D}'$ that can be written as T_f for some locally integrable function f as defined before, is called a regular distribution. Since $\varphi \in \mathcal{D}$, it has compact support, hence T_f is always well-defined.

Because of this, the inner product $\langle f(x), \varphi(x) \rangle$ is equal to $\langle T_f, \varphi \rangle$ for any test function $\varphi \in \mathcal{D}$. As a consequence, T_f is actually linear with respect to f, i.e. $T_{f+g} = T_f + T_g$ and $T_{\lambda \cdot f} = \lambda \cdot T_f$. Therefore, $\langle T_f, \varphi \rangle$ is usually denoted $\langle f, \varphi \rangle$, and this property will prove to be very useful.

Actually, since this notion of a regular distribution will be occurring often throughout the report, it may be useful to recall what integrability actually means. For this, the reader is referred to Appendix A (Definition A.1).

A final example is based on the partial differential operator, as stated in Definition 2.2. This can then immediately be used to give an equivalent definition of a distribution.

Example 3.4. Let $k \in \mathbb{N}^n, a \in \mathbb{R}^n$, then $\langle T, \varphi \rangle \equiv D^k \varphi(a)$ for $\varphi \in \mathcal{D}$ is a distribution.

It will turn out that the continuity condition of Definition 3.1 can be guaranteed whenever the distribution is bounded by partial differential operators. Furthermore, it will be shown in the following proof that this is actually a necessary property of a distribution, and can thus be used to give an equivalent definition.

Proposition 3.5. $T : \mathcal{D} \to \mathbb{C}$ is a distribution if and only if it is linear and for every compact set $K \subset R^n$ there exist constants $C_K > 0$ and $m \in \mathbb{N}$ such that

$$|\langle T, \varphi \rangle| \leq C_K \cdot \sum_{|k| \leq m} \sup |D^k \varphi(x)| \ \ \forall \varphi \in \mathcal{D}(K). \tag{3.1}$$

Note that here, and in future cases, $\sup\limits_{x \in \mathbb{R}^n}$ has been abbreviated to sup.

Proof. Clearly the first two requirements of Definition 3.1 are equivalent to linearity of T. Now for the continuity condition:

"$\Leftarrow$": If T satisfies inequality (3.1) and $\varphi_j \xrightarrow{\mathcal{D}} \varphi$, i.e. $(\varphi_j - \varphi) \xrightarrow{\mathcal{D}} 0$, then clearly $|\langle T, \varphi_j - \varphi \rangle| \xrightarrow{\mathbb{R}} 0$ and thus $\langle T, \varphi_j \rangle \xrightarrow{\mathbb{C}} \langle T, \varphi \rangle$. Hence T is a distribution by Definition 3.1.

"$\Rightarrow$": If T is a distribution, suppose it does not satisfy the proposition. Then there exists a compact set $K \subset \mathbb{R}^n$ such that for all constants $C > 0$ and $m \in \mathbb{N}$ (let us set $C = m \in \mathbb{N}$), there exists a function $\varphi_m \in \mathcal{D}(K)$ such that inequality (3.1) does not hold. Without loss of generality, $\varphi_m \in \mathcal{D}(K)$ can be chosen such that $\langle T, \varphi_m \rangle = 1$ and $|D^k \varphi_m| \leq \frac{1}{m}$ for all $k \in \mathbb{N}^n$ such that $|k| \leq m$. Clearly,

$$\langle T, \varphi_m \rangle = 1 > m \cdot \sum_{|k| \leq m} \sup |D^k \varphi(x)| \ \ \forall m \in \mathbb{N}.$$

So $\varphi_m \xrightarrow{\mathcal{D}} 0$, but $\langle T, \varphi_m \rangle = 1 \xrightarrow{\mathbb{C}} 1$, which contradicts that T is a distribution by Definition 3.1. □

The way in which a distribution can be bounded by this inequality actually tells a lot about the distribution itself, and leads to the following classifications.

Definition 3.6. If T is a distribution such that there exists some $m \in \mathbb{N}$ where T satisfies Proposition 3.5 for all compact sets $K \subset \mathbb{R}^n$, T is said to be of order m, given that m is the smallest such integer. If in addition to this, also a constant $C > 0$ can be chosen independently of K, then T is called a summable distribution.

The latter of these will be used later in this chapter, but the first notion will be used straight away. In order to fully understand the definition, and for later use, let us compute the orders of the distributions that have been introduced so far.

Example 3.7. First consider T_f, as defined in Example 3.3. Then for any compact $K \subset \mathbb{R}^n$,

$$|\langle f, \varphi \rangle| \leq \int_{\mathbb{R}^n} |f(x)\ \varphi(x)|\ dx = \int_K |f(x)\ \varphi(x)|\ dx \leq \|\varphi\|_\infty \cdot \int_K |f(x)|\ dx = \|\varphi\|_\infty \cdot \|f\|_1,$$

for all $\varphi \in \mathcal{D}(K)$, where the L^1- and L^∞-norms are denoted $\|\cdot\|_1$ and $\|\cdot\|_\infty$ respectively (their definitions are recalled in Appendix A, Definition A.1). Since $\|\varphi\|_\infty = \sup |D^0\varphi(x)|$, and $\|f\|_1 \equiv C < \infty$ as f is locally integrable, $|\langle f, \varphi \rangle| \leq C \cdot \sup |D^0\varphi(x)|$, and thus T_f has order 0.

Secondly, consider the Dirac delta $\delta_{(a)}$, as defined in Example 3.2. Clearly, this distribution has order 0, as $|\langle \delta_{(a)}, \varphi \rangle| = |\varphi(a)| \leq \sup |\varphi(x)| = \sup |D^0\varphi(x)|$.

Similarly for the distribution in Example 3.4, $|\langle T, \varphi \rangle| = |D^k\varphi(a)| \leq \sup |D^k\varphi(x)|$, and therefore T has order $|k|$, the same order as D^k.

Finally, note that in the last two cases, inequality (3.1) is shown to hold for the entirety of $\mathbb{R}^n$. It holds in general that if this is the case, then clearly the inequality can be shown to hold for any compact subset $K \subset \mathbb{R}^n$ by choosing the same constant $C > 0$. Since this is also true for the first case, all of these examples are summable distributions by Definition 3.6.

In order to have a better control over the test functions that will be worked with, the following theorem shows the existence of test functions that satisfy specific controllable conditions.

Theorem 3.8. Let $K \subset \mathbb{R}^n$ be compact, then for any open set $O \subset \mathbb{R}^n$ with $K \subset O$, there exists some $\varphi \in \mathcal{D}$ such that $0 \leq \varphi(x) \leq 1\ \forall x \in \mathbb{R}^n$ and $\varphi(x) = 1\ \forall x \in O$.

The proof of this theorem is a very technical one, and the interested reader will therefore be referred to [4] (Lemma 2.4) for its proof. The use of this theorem will become apparent immediately in the following lemma, regarding the regular distribution T_f as introduced in Definition 3.3.

Lemma 3.9. Let f be a continuous function, then $\langle f, \varphi \rangle = 0$ for all $\varphi \in \mathcal{D}$ implies that $f = 0$.

Proof. Assume $f(x_0) \neq 0$ for some $x_0 \in \mathbb{R}^n$, then without loss of generality $\operatorname{Re} f(x_0) > 0$, otherwise take $-f$ or $\pm\imath \cdot f$ instead, such that this is the case. Since f is continuous, there exists a neighborhood $V \subset \mathbb{R}^n$ of x_0 where $\operatorname{Re} f(x) > 0$. Note that since $V \subset \mathbb{R}^n$ is bounded, its closure is compact. Therefore by Theorem 3.8 there exists some $\varphi \in \mathcal{D}$ such that $\varphi(x) = 1\ \forall x \in \overline{V}$ and $0 \leq \varphi(x) \leq 1\ \forall x \in \mathbb{R}^n$. Hence,

$$\operatorname{Re} \langle f, \varphi \rangle = \operatorname{Re} \int_{\mathbb{R}^n} f(x)\ \varphi(x)\ dx > 0.$$

This contradicts the assumption that $\langle f, \varphi \rangle = 0$. Therefore, $f = 0$ must hold. □

Note that this lemma confirms that the final issue of the notation $\langle f, \varphi \rangle$ for T_f is justified, namely that $T_f = T_g \Leftrightarrow \langle f, \varphi \rangle = \langle g, \varphi \rangle \Leftrightarrow \langle f - g, \varphi \rangle = 0 \Leftrightarrow f - g = 0 \Leftrightarrow f = g$, as desired.

3.2. Derivative of a Distribution

Now that the notation for T_f has been fully justified, and the basic properties of distributions have been explored, there are a few more notions that will need to be introduced and fully laid out, in order to apply distributions. All of these will be treated in separate sections, starting this section with a very familiar topic to many: differentiation. In order to grasp what a derivative of a distribution means, let us first consider a special case, namely taking the derivative of the regular distribution T_f, which will be shown to come very naturally.

Example 3.10. Given $f \in \mathcal{E}^1(\mathbb{R})$, then note that by applying partial integration to $T_{f'}$,

$$\langle f', \varphi \rangle = \int_{-\infty}^{\infty} f'(x)\ \varphi(x)\ dx = -\int_{-\infty}^{\infty} f(x)\ \varphi'(x)\ dx = -\langle f, \varphi' \rangle \text{ for } \varphi \in \mathcal{D}.$$

Then in $\mathbb{R}^n$, by partial integration $\langle \frac{\partial f}{\partial x_i}, \varphi \rangle = -\langle f, \frac{\partial \varphi}{\partial x_i} \rangle$ for $\varphi \in \mathcal{D}$. It turns out that there is actually an explicit way to write $\frac{\partial T_f}{\partial x_i} \equiv T_{\frac{\partial f}{\partial x_i}}$, and for this the reader is referred to [5] (Theorem 1.16 and Example 1.17). For now, it is important to note that as derived, $\frac{\partial T_f}{\partial x_i}$ is once again a distribution.

This example can be used as a guess to define differentiation on distributions in general. Throughout the chapter though, this definition will be shown to be a fully justified one.

Definition 3.11. Let $T \in \mathcal{D}'$, then $\langle \frac{\partial T}{\partial x_i}, \varphi \rangle \equiv -\langle T, \frac{\partial \varphi}{\partial x_i} \rangle$ for $\varphi \in \mathcal{D}$.

Note that by this definition, $\frac{\partial T}{\partial x_i}$ is once again a distribution. This also emphasizes the importance of test functions once again. Whereas the classical notion of differentiability does not exist for most distributions, once can still get an idea how such a derivative would act differently upon test functions, using this alternate definition of differentiability.

Example 3.12. As an example of this, consider the Dirac delta $\delta_{(a)}$. Its k^{th} derivative is

$$\langle \frac{d^k \delta_{(a)}}{dx^k}, \varphi \rangle = (-1)^k \cdot \langle \delta_{(a)}, \frac{d^k \varphi}{dx^k} \rangle = (-1)^k \cdot \varphi^{(k)}(a) \text{ for } \varphi \in \mathcal{D}(\mathbb{R}).$$

Then for $\mathbb{R}^n$, $\langle D^k \delta_{(a)}, \varphi \rangle = (-1)^{|k|} \cdot D^k \varphi(a)$ for $\varphi \in \mathcal{D}$ and for all $k \in \mathbb{N}^n$, which by Example 3.4 is indeed a distribution.

However, one can ask oneself, can the partial differential operator as given in Definition 2.2 then also be applied to distributions? By the linearity of the space of distributions $\mathcal{D}'$, as given by Property 3.1, it follows that

$$\langle D^k T, \varphi \rangle = (-1)^{|k|} \cdot \langle T, D^k \varphi \rangle \text{ for } \varphi \in \mathcal{D}.$$

This can be used to define that for any differential operator $D = \sum_{|k| \leq m} a_k \cdot D^k$ (where $a_k \in \mathbb{C}$), the adjoint of D is ${}^t D \equiv \sum_{|k| \leq m} (-1)^{|k|} \cdot a_k \cdot D^k$. By the observation earlier, $\langle DT, \varphi \rangle = \langle T, {}^t D \varphi \rangle$ for $\varphi \in \mathcal{D}$, for any distribution T. Now, let us compute the derivative of the final distribution that has been introduced so far.

Example 3.13. Recall the Heaviside function $H(x) = \begin{cases} 1 \text{ for } x \geq 0, \\ 0 \text{ for } x < 0. \end{cases}$ Then,

$$\langle H', \varphi \rangle = -\langle H, \varphi' \rangle = -\int_0^\infty \varphi'(x)\ dx = \varphi(0) = \langle \delta, \varphi \rangle \text{ for } \varphi \in \mathcal{D}(\mathbb{R}).$$

Hence it can be explicitly expressed that $H' = \delta$.

Now, let us jump to another notion that at first glance might not seem to have to do much with differentiation, but as will be shown in a bit, actually goes hand-in-hand. The notion in question is that of multiplying a distribution with an infinitely differentiable function, yielding a so-called multiplicative distribution.

Definition 3.14. Let $\alpha \in \mathcal{E}, T \in \mathcal{D}'$, then define the multiplicative distribution αT as $\langle \alpha T, \varphi \rangle \equiv \langle T, \alpha \cdot \varphi \rangle$ for $\varphi \in \mathcal{D}$. Note that by Property 2.2, $\alpha \cdot \varphi \in \mathcal{D}$, so αT is indeed a distribution.

Let us first consider an example to explore what this definition means for regular distributions, and then look at a concept that will be critical shorty.

Example 3.15. First note that the multiplicative distribution αT_f is given by

$$\langle \alpha T_f, \varphi \rangle = \langle T_f, \alpha \cdot \varphi \rangle = \langle f, \alpha \cdot \varphi \rangle = \int_{\mathbb{R}^n} f(x)\ \alpha(x)\ \varphi(x)\ dx = \langle \alpha \cdot f, \varphi \rangle = \langle T_{\alpha \cdot f}, \varphi \rangle \text{ for } \varphi \in \mathcal{D}.$$

Hence $\alpha T_f = T_{\alpha \cdot f}$, which again is a regular distribution, as $\alpha \cdot f$ is locally integrable as well. Now let us consider the Dirac delta $\delta_{(a)}$. Then the multiplicative distribution is given by

$$\langle \alpha \delta_{(a)}, \varphi \rangle = \langle \delta_{(a)}, \alpha \cdot \varphi \rangle = \alpha(a) \cdot \varphi(a) = \alpha(a) \cdot \langle \delta_{(a)}, \varphi \rangle \text{ for } \varphi \in \mathcal{D}.$$

Therefore, $\alpha \delta_{(a)} = \alpha(a) \cdot \delta_{(a)}$. For instance $x \delta_{(a)} = a \cdot \delta_{(a)}$, and specifically, $x\delta = 0$.

This last observation is a very important one, as this is a concept that can actually be generalized to all distributions, as can be seen in the next theorem.

Theorem 3.16. Let $T \in \mathcal{D}'(\mathbb{R})$, then if $xT = 0$, $T = c \cdot \delta$ for some $c \in \mathbb{C}$.

Proof. By Theorem 3.8, let $\chi \in \mathcal{D}$ be such that $\chi(x) = 1$ in a neighborhood of 0, $0 \leq \chi(x) \leq 1$ otherwise and fix $\varphi \in \mathcal{D}$. Then set

$$\psi(x) = \begin{cases} \frac{\varphi(x) - \varphi(0) \cdot \chi(x)}{x} & \text{when } x \neq 0, \\ \varphi'(0) & \text{when } x = 0. \end{cases}$$

In the case that $x \neq 0$, $\varphi(x) = \varphi(0) \cdot \chi(x) + x \cdot \psi(x)$, and thus by the assumption that $xT = 0$,

$$\begin{aligned} \langle T, \varphi \rangle &= \langle T, \varphi(0) \cdot \chi + x \cdot \psi \rangle \\ &= \langle \varphi(0) \cdot T, \chi \rangle + \langle xT, \psi \rangle \\ &= \langle T, \chi \rangle \cdot \varphi(0) = c \cdot \langle \delta, \varphi \rangle \text{ for } c = \langle T, \chi \rangle \in \mathbb{C}. \end{aligned}$$ □

Hence it is shown that in the case that $xT = 0$, the distribution is analogous to being constant, namely some multiple of the Dirac delta. But beware not to confuse this with a distribution actually being constant, which by a similar proof is the case whenever $T' = 0$. This final observation also hints to the next logical step, namely showing how to differentiate a multiplicative distribution.

Theorem 3.17. Let $T \in \mathcal{D}', \alpha \in \mathcal{E}$, then $\frac{\partial}{\partial x_i}(\alpha T) = \frac{\partial \alpha}{\partial x_i} T + \alpha \frac{\partial T}{\partial x_i}$.

Proof. Let us restrict the proof to $\mathbb{R}$, as the general proof is almost exactly the same, but involves more notation. First note that for any $\varphi \in \mathcal{D}\mathbb{R}$, $(\alpha \cdot \varphi)' = \alpha' \cdot \varphi + \alpha \cdot \varphi'$, by the product rule. Then by Definitions 3.11 and 3.14, it follows that

$$\begin{aligned}\langle (\alpha T)', \varphi \rangle &= -\langle \alpha T, \varphi' \rangle = -\langle T, \alpha \cdot \varphi' \rangle \\ &= -\langle T, (\alpha \cdot \varphi)' \rangle + \langle T, \alpha' \cdot \varphi \rangle \\ &= \langle T', \alpha \cdot \varphi \rangle + \langle T, \alpha' \cdot \varphi \rangle \\ &= \langle \alpha T', \varphi \rangle + \langle \alpha' T, \varphi \rangle.\end{aligned}$$

Since this holds for any $\varphi \in \mathcal{D}(\mathbb{R})$, $(\alpha T)' = \alpha' T + \alpha T'$. □

3.3. Support of a Distribution

Before moving on to the next topic of interest, it may be useful to realize the peculiarity of the occurrence in Theorem 3.16 again at this point. The notion that a distribution which vanishes when multiplied by x, can be written as a multiple of the Dirac delta, is a property that leads to a very important result at the end of this section. It turns out that almost every distribution can be written as a linear combination of derivatives of Dirac deltas, very comparable to the Taylor expansion of a function and almost of the same form as well.

To understand why this is the case and, most importantly, what class of distributions this can be applied to, it is necessary to first explore supports of distributions. To this end, let us first define the support of a distribution, and note that its definition is very similar to that of functions.

Definition 3.18. Let $T \in \mathcal{D}'$, and O be an open set in $\mathbb{R}^n$ such that $\langle T, \varphi \rangle \neq 0$ for all $\varphi \in \mathcal{D}(O)$. The support of a distribution T, denoted $\operatorname{Supp} T$, is the complement of the largest such open set $O \subset \mathbb{R}^n$.

There are examples of distributions in which this definition can easily be applied. For instance it is not hard to check that $\operatorname{Supp} \delta_{(a)} = \{a\}$ and $\operatorname{Supp} T_f = \operatorname{supp} f$ (try taking any element outside the support and look what happens). Although this looks very equivalent to Definition 2.1, regarding the support for functions, actually finding this set can prove very difficult for more complex distributions. However, recall that by Proposition 3.5, there is an equivalent definition of a distribution using an inequality that will prove very useful in order to bound the support of a distribution.

Proposition 3.19. Let $T \in \mathcal{D}'$ be of finite order $m \in \mathbb{N}$. Then if $\varphi \in \mathcal{D}$ is such that $D^k \varphi(x) = 0$ for all $x \in \operatorname{Supp} T$ and $k \in \mathbb{N}^n$ with $|k| \leq m$, it follows that $\langle T, \varphi \rangle = 0$.

Proof. Let $\varphi \in \mathcal{D}$ be as in the proposition, and let $K \subset \mathbb{R}^n$ be a compact set such that $\operatorname{supp} \varphi \subset K$, then by Proposition 3.5,

$$|\langle T, \varphi \rangle| \leq C_K \cdot \sum_{|k| \leq m} \sup |D^k \varphi(x)| \;\; \forall \varphi \in \mathcal{D}(K) \text{ and for some } C_k > 0.$$

Moreover, by Theorem 3.8, for all $\epsilon > 0$ there exists some $\psi_\epsilon \in \mathcal{D}(K)$ such that $\psi_\epsilon(x) = 1$ for all $x \in K \cap \operatorname{Supp} T$ and $|D^k(\psi_\epsilon \cdot \varphi)| \le \epsilon$ for all $k \in \mathbb{N}^n$ with $|k| \le m$. Note that since $\operatorname{supp} \psi_\epsilon \subset K$,

$$|\langle T, \varphi\rangle| = |\langle T, \psi_\epsilon \cdot \varphi\rangle| \le C_K \cdot \sum_{|k|\le m} \sup |D^k(\psi_\epsilon \cdot \varphi)(x)| \le C_k \cdot (m+1)^n \cdot \epsilon.$$

Since ϵ can be chosen arbitrarily small such that this inequality still holds, $\langle T, \varphi\rangle = 0$. □

This contrapositive of this proposition is most often used, i.e. $\langle T, \varphi\rangle \neq 0$ implies that the proposition does not hold for that test function φ. Concretely, if $\varphi \in \mathcal{D}$ is such that $\langle T, \varphi\rangle \neq 0$ for a distribution T of order $m \in \mathbb{N}^n$, then if $x \in \mathbb{R}^n$ is such that $D^k\varphi(x) \neq 0$ for all $k \in \mathbb{N}^n$ with $|k| \le m$, it is guaranteed that $x \notin \operatorname{Supp} T$.

This can therefore be used to bound the support of a distribution, when choosing the right test functions to apply this to. There is an even stronger consequence of this though, so strong that it deserves its own corollary, while its proof is a trivial insight.

Corollary 3.20. Let $T \in \mathcal{D}'$. If there exists some $\varphi \in \mathcal{D}$ such that $\langle T, \varphi\rangle \neq 0$ and $D^k\varphi(x) \neq 0$ for all $k \in \mathbb{N}^n$ and $x \in \mathbb{R}^n$, then $\operatorname{Supp} T = \{0\}$.

Not only for this section, but throughout the report, a very important characterization of a distribution is whether it has compact support. It turns out that there is an easy and equivalent way of finding such distributions, given in the following proposition.

Proposition 3.21. A distribution T has compact support if and only if there exist a compact set $K \subset \mathbb{R}^n$, $C > 0$ and $m \in \mathbb{N}$ such that for all compact sets $K \subset \mathbb{R}^n$,

$$|\langle T, \varphi\rangle| \le C \cdot \sum_{|k|\le m} \sup |D^k\varphi(x)| \;\; \forall \varphi \in \mathcal{D}(K).$$

Moreover, if this is the case, $\operatorname{Supp} T \subset K$.

Its proof is very similar to that of Proposition 3.19, with the added result that in this case $\operatorname{Supp} T$ is compact. Also note that distributions with compact support are not per se of finite order and summable, as that is only the case if the above inequality holds for all compact sets $K \subset \mathbb{R}^n$.

Using this, the notion of a distribution being constant as in Theorem 3.16 can be taken one step further, by considering what a Taylor expansion of a distribution would look like. First, recall that for $\varphi \in \mathcal{D}(\mathbb{R})$, the m^{th} order Taylor expansion is given by

$$\varphi(x) = \sum_{k=0}^{m} \frac{x^k}{k!} \cdot \varphi^{(k)}(0) + \frac{1}{m!} \cdot \int_0^x (x-t)^m \cdot \varphi^{(m+1)}(t)\; dt.$$

This can be extended to $\mathcal{D}$=$\mathcal{D}(\mathbb{R}^n)$ for any $n \in \mathbb{N}$, giving the Taylor expansion

$$\varphi(x) = \sum_{|k|\le m} \frac{x^k}{k!} \cdot D^k\varphi(0) + \psi(x), \tag{3.2}$$

for some residue function $\psi \in \mathcal{E}$ such that $D^k\psi(0) = 0$ for all $k \in \mathbb{N}^n$ such that $|k| \le m$. Also, recall that $k! = k_1! \cdots k_n!$, and define $x^k \equiv x^{|k|} = x^{k_1} \cdots x^{k_n}$. This can be used to derive a similar Taylor expansion for any distribution T with compact support.

Theorem 3.22. Let T be a distribution with compact support, such that $0 \in \operatorname{Supp} T$. Then

$$T = \sum_{|k| \leq m} c_k \cdot D^k \delta \text{ for some } (c_k) \in \mathbb{C},$$

is the Taylor expansion of T in $\mathcal{D}'$, where m is the order of T, as defined in Definition 3.6.

Proof. Let $\psi \in \mathcal{E}$ be the residue of the m^{th} order Taylor expansion of $\varphi \in \mathcal{D}$, as in equation (3.2). Then $D^k\psi(0) = 0$ for all $k \in \mathbb{N}^n$ such that $|k| \leq m$, and all its derivatives vanish on the support of T. This can easily be computed by taking the derivatives of equation (3.2). Then by Property 3.19, $\langle T, \psi \rangle = 0$. Hence,

$$\begin{aligned}
\langle T, \varphi \rangle &= \sum_{|k| \leq m} \langle T, \frac{x^k}{k!} \cdot D^k\varphi(0) \rangle + \langle T, \psi \rangle = \sum_{|k| \leq m} \langle T, \frac{x^k}{k!} \cdot D^k\varphi(0) \rangle \\
&= \sum_{|k| \leq m} \langle T, \frac{x^k}{k!} \cdot \langle \delta, D^k\varphi \rangle \rangle = \sum_{|k| \leq m} \langle T, (-1)^{|k|} \cdot \frac{x^k}{k!} \cdot \langle D^k\delta, \varphi \rangle \rangle \\
&= \sum_{|k| \leq m} \langle T, (-1)^{|k|} \cdot \frac{x^k}{k!} \rangle \cdot \langle D^k\delta, \varphi \rangle.
\end{aligned}$$

This implies that

$$T = \sum_{|k| \leq m} c_k \cdot D^k\delta, \text{ where } c_k = \langle T, (-1)^{|k|} \cdot \frac{x^k}{k!} \rangle \in \mathbb{C}. \qquad \square$$

3.4. Extension of a Distribution

In the proof of Theorem 3.22 last section, the attentive reader may have noticed that one major detail was glanced over. Property 3.19 was used to justify that $\langle T, \psi \rangle = 0$, and while that will turn out to be correct, how does one even interpret this expression? After all, ψ was chosen to be a C^∞ function only, not necessarily a test function, so how to apply a distribution to it? It turns out that any distribution with compact support can be extended to a linear continuous map in $\mathcal{E}'$ (and vice versa), the proof of which will be the goal of this section. To do this however, it is crucial to understand what the exact relationship is between $\mathcal{D}$ and $\mathcal{E}$, and for that it is necessary to define two more notions of convergence.

Definition 3.23. A sequence $(T_j) \in \mathcal{D}'$ is said to converge to $T \in \mathcal{D}'$ if $\langle T_j, \varphi \rangle \xrightarrow{\mathbb{C}} \langle T, \varphi \rangle$ for all $\varphi \in \mathcal{D}$. This is denoted $T_j \xrightarrow{\mathcal{D}'} T$.

An easily verifiable consequence of this is that if $T_j \xrightarrow{\mathcal{D}'} T$, then $D^kT_j \xrightarrow{\mathcal{D}'} D^kT$ for all $k \in \mathbb{N}^n$. To see this, fix $k \in \mathbb{N}^n$, then

$$\langle D^kT_j, \varphi \rangle = (-1)^{|k|} \cdot \langle T_j, D^k\varphi \rangle \xrightarrow{\mathbb{C}} (-1)^{|k|} \cdot \langle T, D^k\varphi \rangle = \langle D^kT, \varphi \rangle \ \forall \varphi \in \mathcal{D}.$$

Definition 3.24. A sequence $(\varphi_j) \in \mathcal{E}$ is said to converge to $\varphi \in \mathcal{E}$ if $D^k\varphi_j \xrightarrow{\mathcal{D}(K)} D^k\varphi$ for any compact set $K \subset \mathbb{R}^n$ and for all $k \in \mathbb{N}^n$. This is denoted $\varphi_j \xrightarrow{\mathcal{E}} \varphi$.

Since $\mathcal{D} \subset \mathcal{E}$, it is necessary to check if this definition is consistent with Definition 2.4. To this end, note that if $\varphi_j \xrightarrow{\mathcal{D}} \varphi$, by Corollary 2.5 $D^k\varphi_j \xrightarrow{\mathcal{D}} D^k\varphi$ for all $k \in \mathbb{N}^n$. Hence $D^k\varphi_j \xrightarrow{\mathcal{D}(K)} D^k\varphi$

for any compact set $K \subset \mathbb{R}^n$ and for all $k \in \mathbb{N}^n$, and by the last definition, $\varphi_j \xrightarrow{\mathcal{E}} \varphi$. Thus, this definition is indeed consistent with convergence on $\mathcal{D}$. Now all the necessary tools are in place to elaborate on the relation $\mathcal{D} \subset \mathcal{E}$ further.

Property 3.2. $\mathcal{D}$ is dense in $\mathcal{E}$.

Proof. Recall that a closed ball around the origin is given by $\overline{B_j(0)} = \{x \in \mathbb{R}^n \mid \|x\| \leq j\}$, where in this case the radius is some $j \in \mathbb{N}$. Fix $\varphi \in \mathcal{E}$, and note that by Theorem 3.8, $\alpha_j \in \mathcal{D}$ can be chosen such that $0 \leq \alpha_j(x) \leq 1 \ \forall x \in \mathbb{R}^n$ and $\alpha_j(x) = 1 \ \forall x \in \overline{B_j(0)}$, for any $j \in \mathbb{N}$. By Property 2.2, $(\alpha_j \cdot \varphi) \in \mathcal{D}$, and since $\alpha_j(x) = 1 \ \forall x \in \overline{B_j(0)}$, $D^k(\alpha_j \cdot \varphi) \xrightarrow{\mathcal{D}(\overline{B_j(0)})} D^k\varphi$ for all $k \in \mathbb{N}^n$. Then by Definition 3.24, $\alpha_j \cdot \varphi \xrightarrow{\mathcal{E}} \varphi$, hence $\varphi \in \mathcal{D}$ and thus $\mathcal{D}$ is dense in $\mathcal{E}$. □

Now, $\mathcal{E}'$ can be defined analogously to $\mathcal{D}'$ in Definition 3.1, by simply adjusting its continuity requirement.

Definition 3.25. The map $L : \mathcal{E} \to \mathbb{C}$ is an element of $\mathcal{E}'$ if for all $\varphi, \psi \in \mathcal{D}$ and $\lambda \in \mathbb{C}$:

1. $L(\varphi + \psi) = L(\varphi) + L(\psi)$
2. $L(\lambda \cdot \varphi) = \lambda \cdot L(\varphi)$
3. If $\varphi_j \xrightarrow{\mathcal{E}} \varphi$, then $L(\varphi_j) \xrightarrow{\mathbb{C}} L(\varphi)$

These continuous linear forms are very similar to the notions of distributions and therefore $L(\varphi)$ will also be denoted $\langle L, \varphi \rangle$. Adding to this the previous observation that $\mathcal{D}$ is dense in $\mathcal{E}$, the theorem this section has been working to should not come as a surprise.

Theorem 3.26.

1. Every distribution with compact support can be uniquely extended to a continuous linear form on $\mathcal{E}$.
2. The restriction of any continuous linear form on $\mathcal{E}$ to $\mathcal{D}$ is a distribution with compact support.

Proof.

1. Let $T \in \mathcal{D}'$ have compact support, then by Theorem 3.8 there exists some $\alpha \in \mathcal{D}$ such that $0 \leq \alpha(x) \leq 1 \ \forall x \in \mathbb{R}^n$ and $\alpha(x) = 1 \ \forall x \in \operatorname{Supp} T$. Now it is possible to let T act on $\alpha \cdot \varphi$ for any $\varphi \in \mathcal{E}$, since $\varphi|_{\operatorname{Supp} T} \in \mathcal{D}$ and outside $\operatorname{Supp} T$, T vanishes. Then $\langle T, \alpha \cdot \varphi \rangle = \langle T, \varphi \rangle$ on $\operatorname{Supp} T$, so let $\langle L, \varphi \rangle \equiv \langle T, \alpha \cdot \varphi \rangle$ for any $\varphi \in \mathcal{E}$. Then since T is continuous and linear on $\mathcal{D}$, and $L|_{\mathcal{D}} = T$, $L \in \mathcal{E}$. Thus, since $\mathcal{D}$ is dense in $\mathcal{E}$, L is the unique extension of T to $\mathcal{E}'$.
2. Clearly, the restriction $T \equiv L|_{\mathcal{D}} \in \mathcal{D}'$ for any $L \in \mathcal{E}'$, by continuity and linearity. But suppose T does not have compact support, then for all $m \in \mathbb{N}$ there exists some $\varphi_m \in \mathcal{D}$ such that $\operatorname{supp} \varphi_m \cap B_m(0) = \emptyset$ and $\langle T, \varphi_m \rangle = 1$, where for the latter part Theorem 3.8 was used. Note that due to the prior statement, $\varphi_m \xrightarrow{\mathcal{E}} 0$, so since $L \in \mathcal{E}$, $L(\varphi_m) \xrightarrow{\mathbb{C}} L(0) = 0$ by Definition 3.25. However this contradicts that $\langle L, \varphi_m \rangle = \langle T, \varphi_m \rangle = 1$ for all $m \in \mathbb{N}$. Thus, the restriction T of any $L \in \mathcal{E}'$ must have compact support. □

4. Convolution

4.1. Tensor Products

Last chapter, distributions have been defined and many of its properties were introduced. Furthermore, several notions regarding distributions have been familiarized, here and there a few insightful applications have been mentioned, but it is now time to start working towards a whole field of applications. However, to understand how to express an application in terms of distributions, one needs to understand convolution. At the end of the chapter it will be possible to convert a differential equation to a convolution equation and understand how to solve this, but let us start with a more familiar concept that underlies this: the tensor product.

In order to define how the tensor product acts upon distributions, let us first recall what it looks like on functions, and use that to extend it to distributions. As a final note, let us define for this entire chapter that the sets $X, Y \subset \mathbb{R}^n$ and that $x \in X$, $y \in Y$.

Definition 4.1. Let $f : X \to \mathbb{C}$, $g : Y \to \mathbb{C}$, then the tensor product $f \otimes g : X \times Y \to \mathbb{C}$ is defined by $(f \otimes g)(x, y) \equiv f(x) \cdot g(y)$.

Note that whenever both functions are locally integrable, the tensor product $f \otimes g$ is also locally integrable. In this case, for $u \in \mathcal{D}(X)$, $v \in \mathcal{D}(Y)$,

$$\begin{aligned}
\langle f \otimes g, u \otimes v \rangle &= \int_X \int_Y (f \otimes g)(x, y) \cdot (u \otimes v)(x, y)\ dy\ dx \\
&= \int_X \int_Y f(x)\ g(y)\ u(x)\ v(y)\ dy\ dx \\
&= \int_X f(x)\ u(x)\ dx \cdot \int_Y g(y)\ v(y)\ dy \\
&= \langle f, u \rangle \cdot \langle g, v \rangle.
\end{aligned}$$

However, how does this regular distribution act upon a test function that is not per say a tensor product of two test functions? Let $\varphi \in \mathcal{D}(X \times Y)$, then by Fubini's theorem,

$$\begin{aligned}
\langle f \otimes g, \varphi \rangle &= \int_X \int_Y (f \otimes g)(x, y) \cdot \varphi(x, y)\ dy\ dx \\
&= \int_X f(x) \int_Y g(y)\ \varphi(x, y)\ dy\ dx = \langle f, \langle g, \varphi \rangle \rangle \\
&= \int_Y g(y) \int_X f(x)\ \varphi(x, y)\ dx\ dy = \langle g, \langle f, \varphi \rangle \rangle.
\end{aligned}$$

This observation is used to find a distribution that acts upon test functions exactly as the tensor product of two distributions would.

Proposition 4.2. Let $S \in \mathcal{D}'(X)$, $T \in \mathcal{D}'(Y)$, then there exists a unique $W \in \mathcal{D}'(X \times Y)$ such that $\langle W, u \otimes v \rangle = \langle S, u \rangle \cdot \langle T, v \rangle$ for all $u \in \mathcal{D}(X)$ and $v \in \mathcal{D}(Y)$. Then W is called the tensor product of S and T and denoted $W = S \otimes T$.

Its existence is not very difficult to see, but since the proof of that and the uniqueness thereof is very technical, the interested reader is redirected to [3] (Theorem 3.8) for this.

Note that from this proposition, the linearity and associativity of the tensor product of distributions follows. The examples with regular distributions before allow for a well-educated guess on how the tensor product acts on test functions that are not necessarily tensor products themselves. And the reader would be right in making such guess, as there is actually a theorem analogous to that of Fubini's theorem for distributions. To understand why, however, let us first introduce two lemmas that help to understand what $\mathcal{D}(X \times Y)$ looks like and to define a special test function on it.

Lemma 4.3. $\mathcal{D}(X) \otimes \mathcal{D}(Y) \equiv \{u \otimes v \mid u \in \mathcal{D}(X), v \in \mathcal{D}(Y)\}$ is dense in $\mathcal{D}(X \times Y)$.

Lemma 4.4.

1. If $\varphi \in \mathcal{D}(X \times Y)$, then $\chi(x) \equiv \langle T, \varphi(x, y)\rangle \in \mathcal{D}(X)$ for every $T \in \mathcal{D}'(Y)$.
2. If $\varphi \in \mathcal{E}(X \times Y)$, then for every $T \in \mathcal{D}'(Y)$ there exists an extension $L \in \mathcal{E}'(Y)$ such that $\chi(x) \equiv \langle L, \varphi(x, y)\rangle \in \mathcal{E}(X)$.

The content of both lemmas should not be very surprising, once their exact meaning have been grasped. Nonetheless, their proofs rely on very technical details, thus the interested reader is redirected to [3] (Lemma 3.7 and Corollary 3.4 respectively) for their proofs.

Only Lemma 4.4 may offer a surprising addition, as from this it follows that everything holding below for a distribution with compact support, will also hold for its extension in $\mathcal{E}'$. This will be useful to keep in mind for later this chapter, but let us first prove that distributions satisfy a property, that is very analogous to Fubini's theorem.

Property 4.1. Let $S \in \mathcal{D}'(X)$, $T \in \mathcal{D}'(Y)$ and fix $\varphi \in \mathcal{D}(X \times Y)$. Then it follows that $\langle S \otimes T, \varphi\rangle = \langle S, \langle T, \varphi\rangle\rangle = \langle T, \langle S, \varphi\rangle\rangle$.

Proof. Let us prove the first equality, then the second one follows from exchanging X and Y. First note that by Lemma 4.4, $\langle S, \langle T, \varphi\rangle\rangle = \langle S, \chi\rangle$ for $\chi \in \mathcal{D}(X)$, and thus the desired equality is well-defined. Moreover, since by Lemma 4.3 $\mathcal{D}(X) \otimes \mathcal{D}(Y)$ is dense in $\mathcal{D}(X \times Y)$, for all $\varphi \in \mathcal{D}(X \times Y)$ there exist sequences $(u_j) \in \mathcal{D}(X), (v_j) \in \mathcal{D}(Y)$, such that

$$\varphi(x, y) = \sum_{j=1}^{m} u_j(x) \otimes v_j(y)$$

for some $m \in \mathbb{N}$, possibly equal to infinity. Usually the latter case is a problem, but the linearity is the only property necessary here, as then

$$\begin{aligned}
\langle S \otimes T, \varphi\rangle &= \langle S \otimes T, \sum_{j=1}^{m} u_j \otimes v_j\rangle = \sum_{j=1}^{m} \langle W, u_j \otimes v_j\rangle \\
&= \sum_{j=1}^{m} \langle S, u_j\rangle \cdot \langle T, v_j\rangle = \langle S, \sum_{j=1}^{m} \langle T, v_j\rangle \cdot u_j\rangle \\
&= \langle S, \langle T, \sum_{j=1}^{m} u_j \otimes v_j\rangle\rangle = \langle S, \langle T, \varphi\rangle\rangle,
\end{aligned}$$

where the second line follows from Proposition 4.2. □

To further illustrate what a tensor product looks like explicitly, let us treat two examples of distributions encountered before, that will be useful to have computed for later use.

Example 4.5. Let $\delta_{(a)} \in \mathcal{D}'(X)$ and $\delta_{(b)} \in \mathcal{D}'(Y)$ be the Dirac deltas as before. Then $\delta_{(a)} \otimes \delta_{(b)} = \delta_{(a,b)}$, which is in $\mathcal{D}'(X \times Y)$ and thus acts as expected: $\langle \delta_{(a,b)}, \varphi \rangle = \varphi(a,b)$ for $\varphi \in \mathcal{D}(X \times Y)$.

Next, fix $k, l \in \mathbb{N}^n$ and let D_x^k denote the partial differential operator acting on $x \in X$. Then $D_x^k D_y^l (S \otimes T) = D_x^k S \otimes D_y^l T$ is how differential operators act on tensor products.

4.2. Convolution and Regularization

Now, the tensor product for distributions has been defined. In order to come up with Proposition 4.2, first the tensor product for functions was considered, as stated in Definition 4.1. For the notion of convolution, let us walk through a similar process, and thus first recall what this looks like for functions. To this end, suppose $f \in L^1(X)$ and $g \in L^1(Y)$, then their convolution product is defined as

$$(f * g)(x) \equiv \int_{-\infty}^{\infty} f(x-y)\ g(y)\ dy = \int_{-\infty}^{\infty} f(y)\ g(x-y)\ dy\ a.e. \tag{4.1}$$

Here a.e. stands for 'almost everywhere' and is a notion from measure theory. The reason that this last equation is not a formal definition, is because the domain of $f * g$ is not obvious. To this end, let us now immediately define the convolution product for distributions, where this will be formally stated as well, and thereafter consider why (or actually when) this is well-defined.

Definition 4.6. Let $S \in \mathcal{D}'$, $T \in \mathcal{D}'$, then the convolution product $S * T$ is defined as

$$\begin{aligned} \langle S * T, \varphi(z) \rangle &\equiv \langle S \otimes T, \varphi(x+y) \rangle \text{ for } \varphi \in \mathcal{D}(\mathbb{R}^n) \\ &= \langle S \otimes T, \psi(x,y) \rangle \text{ for } \psi(x,y) \equiv \varphi(x+y) \in \mathcal{E}(X \times Y). \end{aligned}$$

This may be confusing, so let us zoom in on the details. As seen before, the tensor product $S \otimes T$ can be extended to act upon a function $\psi(x,y) \in \mathcal{E}(X \times Y)$. However the convolution product $S * T$ is defined such that it chooses that function $\psi(x,y)$ to be $\varphi(x+y)$, and lets the tensor product $S \otimes T$ act on it. Therefore, the convolution product $S * T$ acts upon a test function $\varphi(z) \in \mathcal{D}(\mathbb{R}^n)$, and thus $(S * T) \in \mathcal{D}'$. This is crucial to understand, but when done so, it is clear that the convolution product is both linear and associative, since the tensor product is so as well.

Note that since $\varphi \in \mathcal{D}$, it has compact support, but $\psi(x,y) = \varphi(x+y)$ does not (except when $\varphi = 0$), which is why $\psi(x,y) \in \mathcal{E}(X \times Y)$. From the note under Lemma 4.4 it follows that the evaluation of the tensor product on ψ, $\langle S \otimes T, \psi(x,y) \rangle$, is only well-defined whenever either S or T has compact support. Actually, this is merely a special case, and it can be deduced that it exists whenever $\operatorname{Supp}(S * T) \cap \operatorname{supp} \psi$ is compact. However, note that $\psi(x,y) = 0 \Leftrightarrow \varphi(x+y) = 0 \Leftrightarrow x + y = 0$, since $\varphi \in \mathcal{D}$. The conclusion of this is stated in the following theorem.

Theorem 4.7 (Convolution condition)**.** Let $S \in \mathcal{D}'(X)$, $T \in \mathcal{D}'(Y)$, then if for every compact set $K \subset \mathbb{R}^n$. the set $\{(x,y) \mid (x+y) \in K,\ x \in \operatorname{Supp} S,\ y \in \operatorname{Supp} T\}$ is compact, $S * T$ is well-defined for all $\varphi \in \mathcal{D}$.

Moreover, it follows that in this case $S*T$ is commutative. This theorem is called the convolution condition, and S or T having compact support is a special case for which it always holds. There are in fact two other, more specific, cases for which the convolution condition is satisfied.

Example 4.8. In $\mathbb{R}$, when both $\operatorname{Supp} S$ and $\operatorname{Supp} T$ are bounded from either the left or right, $S * T$ exists. This is often used to work out one-dimensional differential equations, on which convolutions can be naturally defined that satisfy this property.

In $\mathbb{R}^4$, let $\mathbb{H}_4 \equiv \{(t, x, y, z) \mid t \geq 0\}$ be the positive half-space and $\mathbb{E}_4 \equiv \{(t, x, y, z) \mid t \geq 0,\ t^2 - x^2 - y^2 - z^2 = 0\}$ be the positive light cone. Then $S * T$ exists whenever both $\operatorname{Supp} S \subset \mathbb{H}_4$ and $\operatorname{Supp} T \subset \mathbb{E}_4$. This is often used when solving differential equations in space-time.

Since the definition for convolution of distributions is very similar to the convolution product for functions, it will not be difficult to compute the convolution product of two regular convolutions, and observe that this is almost exactly equal to equation (4.1).

Theorem 4.9. Let $T_f, T_g \in \mathcal{D}'$ satisfy the convolution condition. Then

$$(T_f * T_g)(x) = \int_{\mathbb{R}^n} f(x-y)\ g(y)\ dy = \int_{\mathbb{R}^n} f(y)\ g(x-y)\ dy \ a.e.$$

Furthermore, $T_f * T_g$ is a locally integrable function itself, and thus $T_f * T_g = T_{f*g}$.

Proof. Note that for any $\varphi \in \mathcal{D}$

$$\begin{aligned}
\langle T_f * T_g, \varphi \rangle = \langle T_f \otimes T_g, \varphi(x+y) \rangle &= \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} f(x)\ g(y)\ \varphi(x+y)\ dx\ dy \\
&= \int_{\mathbb{R}^n} g(y) \int_{\mathbb{R}^n} f(x)\ \varphi(x+y)\ dx\ dy \\
&= \int_{\mathbb{R}^n} g(y) \int_{\mathbb{R}^n} f(x-y)\ \varphi(x)\ dx\ dy \ a.e. \\
&= \int_{\mathbb{R}^n} \varphi(x) \int_{\mathbb{R}^n} f(x-y)\ g(y)\ dy\ dx \ a.e. \\
&= \langle \int_{\mathbb{R}^n} f(x-y)\ g(y)\ dy, \varphi \rangle \ a.e.
\end{aligned}$$

where the second and fourth line follow from Fubini's theorem. Hence

$$T_f * T_g = \int_{\mathbb{R}^n} f(x-y) \cdot g(y)\ dy \ a.e. \text{ and } \langle T_f * T_g, \varphi \rangle = \langle f * g, \varphi \rangle, \text{ so } T_f * T_g = T_{f*g}$$

A similar proof holds for the symmetrical case. □

Now, the result of this theorem may be a remarkable one; namely that the convolution product of a regular distribution is once again a regular distribution - while by its definition it is a tensor product acting upon a special function. Yet on the other hand, one could have expected this result, for the convolution product is also classically defined on functions and therefore the resulting product should be a function as well.

This opens the door to the question whether it would be possible to do the same thing for other, non-regular, distributions. As the next theorem will show, it is indeed the case that the convolution of any distribution with a continuous function will yield a regular distribution, an operation that is called regularization.

Theorem 4.10. Let $\alpha \in \mathcal{E}$ and $T \in \mathcal{D}'$, then if $T * T_\alpha$ exists, it is equal to T_f, for $f(x) \equiv \langle T, \alpha(x-y)\rangle \in \mathcal{E}$. Then T_f is denoted $T * \alpha$, and called the regularization of T by α.

Proof. Let us treat the case that T and T_α satisfy the convolution condition. Recall that by the second part of Lemma 4.4, $f(x) \equiv \langle T, \alpha(x-y)\rangle \in \mathcal{E}$, as $\alpha \in \mathcal{E}$. Then

$$\begin{aligned}
\langle T * T_\alpha, \varphi\rangle &= \langle T \otimes T_\alpha, \varphi(x+y)\rangle = \langle T, \langle \alpha, \varphi(x+y)\rangle\rangle \\
&= \langle T, \int_{\mathbb{R}^n} \alpha(x)\ \varphi(x-y)\ dx\rangle = \langle T, \int_{\mathbb{R}^n} \alpha(x-y)\ \varphi(x)\ dx\rangle \\
&= \langle T, \langle \varphi, \alpha(x-y)\rangle\rangle = \langle T \otimes \varphi, \alpha(x-y)\rangle \\
&= \langle \varphi, \langle T, \alpha(x-y)\rangle\rangle = \int_{\mathbb{R}^n} \varphi(x) \cdot \langle T, \alpha(x-y)\rangle\ dx \\
&= \int_{\mathbb{R}^n} f(x)\ \varphi(x)\ dx = \langle f, \varphi\rangle \text{ for } \varphi \in \mathcal{D},
\end{aligned}$$

where the first, third and fourth line were obtained by using Property 4.1. Hence $T*T_\alpha = T_f$. □

Let us remember the shortcut that by Property 4.1,

$$\langle T * T_\alpha, \varphi\rangle = \langle T, \langle \alpha, \varphi(x+y)\rangle\rangle = \langle\langle T, \alpha(x-y)\rangle, \varphi\rangle \text{ for } \varphi \in \mathcal{D}. \tag{4.2}$$

Here the first equation holds for any convolution product, and both hold when the conditions of Theorem 4.10 apply.

Apart from regularization, there are some other special convolutions, most of which will also prove helpful later on.

Example 4.11. Note that for all $T \in \mathcal{D}'$, by equation (4.2),

$$\langle T * \delta, \varphi\rangle = \langle T, \langle \delta, \varphi(x+y)\rangle\rangle = \langle T, \varphi(x)\rangle \text{ for } \varphi \in \mathcal{D}.$$

Hence it follows that $T * \delta = T$ for any distribution T.

A more general result can be shown as well. Given a distribution $T \in \mathcal{D}'$, let its translation be denoted $\tau_{(a)}T$ and defined as $\langle \tau_{(a)}T, \varphi(x)\rangle \equiv \langle T, \varphi(x+a)\rangle$ for $\varphi \in \mathcal{D}$. Then by equation (4.2),

$$\langle T * \delta_{(a)}, \varphi\rangle = \langle T, \langle \delta_{(a)}, \varphi(x+y)\rangle = \langle T, \varphi(x+a)\rangle = \langle \tau_{(a)}T, \varphi\rangle \text{ for } \varphi \in \mathcal{D}.$$

Therefore $T * \delta_{(a)} = \tau_{(a)}T$ for any distribution T, which is consistent with the case that $a = 0$.

The following example is a crucial one, tying convolution to differentiation, a result that will be used elaborately throughout the remaining chapters. Note that in $\mathbb{R}$, by equation (4.2),

$$\langle \delta^{(m)} * T, \varphi\rangle = \langle T, \langle \delta^{(m)}, \varphi(x+y)\rangle\rangle = \langle T, (-1)^m \cdot \varphi^{(m)}(x)\rangle = \langle \frac{d^m}{dx^m}T, \varphi\rangle \text{ for } \varphi \in \mathcal{D}.$$

Then in $\mathbb{R}^n$, it follows that $\frac{\partial^m}{\partial x_i^m}\delta * T = \frac{\partial^m}{\partial x_i^m}T$. As a consequence, $D\delta * T = DT$ for any differential operator D.

Hence once may now see how to turn differentiation of a distribution into a convolution product. In order to find out what differentiation of a convolution looks like, it is first necessary to expand the convolution product to multiple distributions. This will actually allows us to kill two birds with one stone, as it then also generalizes the convolution product itself, a result that can be of good use later on.

Proposition 4.12. Let $T_1, \dots, T_m \in \mathcal{D}'$, then $\langle T_1 * \ldots * T_m, \varphi \rangle \equiv \langle T_1 \otimes \ldots \otimes T_m, \varphi(x_1 + \cdots + x_m) \rangle$ for $\varphi \in \mathcal{D}$ exists, is associative and commutative if either:

1. All, or all but one, distributions have compact support.

2. In the case that $\mathbb{R}^n = \mathbb{R}$, all supports are bounded from either the left or right.

3. In the case that $\mathbb{R}^n = \mathbb{R}^4$, all distributions have their support contained in $\mathbb{H}_4$, and at least one has its support contained in $\mathbb{E}_4$.

The proof of this is simply an extension of Example 4.8 and the note above it.

Theorem 4.13. Let $S, T \in \mathcal{D}'$ be such that $S * T$ exists, then:

1. $\frac{\partial}{\partial x_i}(S * T) = \frac{\partial}{\partial x_i} S * T = S * \frac{\partial}{\partial x_i} T$.

2. $\tau_{(a)}(S * T) = \tau_{(a)} S * T = S * \tau_{(a)} T$.

Proof. Both of these follow from the observations in Example 4.11 and commutativity:

1. $\frac{\partial}{\partial x_i}(S * T) = \frac{\partial}{\partial x_i} \delta * S * T = \frac{\partial}{\partial x_i} S * T = S * \frac{\partial}{\partial x_i} T$.

2. $\tau_{(a)}(S * T) = \delta_{(a)} * S * T = \tau_{(a)} S * T = S * \tau_{(a)} T$. □

Note that by repeated application of this theorem, it holds for the convolution of any number of distributions, and for any order of differentiation.

4.3. Convolution Equations

Now all tools are in place to go back to the goal at the start of this chapter, namely expressing differential equations in terms of convolution equations. As all prerequisites have already been defined, one may have an educated guess on how this is to be done. However, in order to compute the step after that, solving the resulting equation, it would be wise to first explore when convolution equations can be solved. Fortunately, this turns out to be analogous to working this out in linear algebra. This will therefore briefly be treated first, but let us make sure not to skip over some of the technicalities involved.

Definition 4.14. $\mathcal{A}' \subset \mathcal{D}'$ is called a convolution algebra if:

1. $S, T \in \mathcal{A}' \Rightarrow S * T \in \mathcal{A}'$.

2. $\delta \in \mathcal{A}'$.

3. The convolution product " $*$ " on $\mathcal{A}'$ is associative and commutative.

Note that this is similar to the definition of an algebra in linear algebra; namely a vector space, closed under some associative operation.

Example 4.15. The following examples of convolution algebras are used often, since on these the convolution product is guaranteed to exist by Proposition 4.12:

- $\mathcal{E}'$, the operators that are extensions of distributions with compact support.
- $\mathcal{D}'_+ \equiv \mathcal{D}'([0,\infty\rangle)$ and $\mathcal{D}'_- \equiv \mathcal{D}'(\langle-\infty,0])$, which are both subsets of $\mathcal{D}'(\mathbb{R})$ in which all distributions have their support bounded from either the left or right side respectively.
- $\mathcal{D}'(\mathbb{E}_4)$, which are the distributions having their support entirely contained within $\mathbb{E}_4$, forming a special case of that in Proposition 4.12.

Having mentioned subspaces of distributions on which the convolution product is always defined, it is now possible to state when a convolution equation can be solved, similarly as in linear algebra.

Theorem 4.16. Given any $T \in \mathcal{A}'$, the convolution equation $S * X = T$ is solvable for $X \in \mathcal{A}'$ if and only if $S \in \mathcal{A}'$ has an inverse in $\mathcal{A}'$, i.e. there exists an inverse $S^{-1} \in \mathcal{A}'$ such that $S * S^{-1} = S^{-1} * S = \delta$. In that case, the unique solution is given by $X = S^{-1} * T$.

As stated, the proof of this is similar to any proof provided in linear algebra.

This raises the question though; which distributions have inverses? A simple but disappointing observation is that no regular distribution $T_f \in \mathcal{A}'$ with $f \in \mathcal{E}$ has an inverse in $\mathcal{A}'$. Assume that this is the case, i.e. $T_f^{-1} \in \mathcal{A}'$, then it follows from Theorem 4.10 that by regularization $T_f * T_f^{-1} = T_g$ for some $g \in \mathcal{E}$. On the other hand, by Theorem 4.16 $T_f * T_f^{-1} = \delta \notin \mathcal{E}$, which leads to a contradiction. Like in linear algebra, the following corollary holds for convolution algebras.

Corollary 4.17. If $S, T \in \mathcal{A}'$ are invertible, then $S*T \in \mathcal{A}'$ is so too, and $(S*T)^{-1} = T^{-1}*S^{-1}$.

A crucially important result of Theorem 4.16 however, is that derivatives of δ have inverses in $\mathcal{D}'_+$ and $\mathcal{D}'_-$. This will open the door to actually being able to convert differential equations into convolution equations. In this light, only the case in $\mathcal{D}'_+$ will be treated below and the rest of the chapter, but keep in mind that all of that holds in $\mathcal{D}'_-$ as well.

Theorem 4.18. In $\mathcal{D}'_+$, fix $m \in \mathbb{N}$ and consider the m^{th} order differential operator, given by

$$D = \sum_{k=0}^{m} a_k \cdot \frac{d^k}{dx^k} \text{ with } a_m = 1. \tag{4.3}$$

Then $D\delta$ is invertible in $\mathcal{D}'_+$ and $(D\delta)^{-1} = H \cdot Z$, where $H(x)$ is the Heaviside function and $Z \in \mathcal{E}$ is the solution of $DZ = 0$, with initial conditions $Z^{(k)}(0) = 0$ for all $0 \leq k \leq m-2$ and $Z^{(m-1)}(0) = 1$.

Proof. Clearly, there exists a unique solution to $DZ = 0$ with the given initial conditions. Recall that $H' = \delta$, then

$$\begin{aligned}
(H \cdot Z)' &= H \cdot Z' + Z \cdot \delta = H \cdot Z' + Z(0) \cdot \delta = H \cdot Z' \\
(H \cdot Z)'' &= (H \cdot Z')' = H \cdot Z'' + Z'(0) \cdot \delta = H \cdot Z'' \\
&\vdots \\
(H \cdot Z)^{(m-1)} &= (H \cdot Z^{(m-2)})' = H \cdot Z^{(m-1)} + Z^{(m-2)}(0) \cdot \delta = H \cdot Z^{(m-1)} \\
(H \cdot Z)^{(m)} &= (H \cdot Z^{(m-1)})' = H \cdot Z^{(m)} + Z^{(m-1)}(0) \cdot \delta = H \cdot Z^{(m)} + \delta.
\end{aligned}$$

Hence $D(H \cdot Z) = H \cdot DZ + \delta = \delta$, as $DZ = 0$. So $D\delta * (H \cdot Z) = D(H \cdot Z) = \delta$. □

Note that the notation used here is a bit sloppy, as the resulting inverse is actually the regular distribution $T_{H\cdot Z}$, but let us use this notation instead to keep the equations clean. Overall, this procedure may look complicated at first sight, but usually in applications the differential equation for Z is a very recognizable one. To illustrate this, let us compute the inverse of a special differential operator applied to δ.

Example 4.19. Let $D = \frac{d}{dx} - \lambda$, then $D\delta = \delta' - \lambda \cdot \delta$. By Theorem 4.18, $(\delta' - \lambda \cdot \delta)^{-1} = H \cdot Z$, for $Z \in \mathcal{E}$ satisfying

$$DZ = Z' - \lambda \cdot Z = 0 \text{ with } z(0) = 1.$$

Clearly, $Z(x) = e^{\lambda x}$ is the unique solution, and thus $(\delta' - \lambda \cdot \delta)^{-1} = H(x) \cdot e^{\lambda x}$.

Similarly, this result can be extended by noting that for $D = (\frac{d}{dx} - \lambda)^m$, $D\delta = (\delta' - \lambda \cdot \delta)^m$, and by a computation very similar to this one, it follows that

$$(\delta' - \lambda \cdot \delta)^{m^{-1}} = H(x) \cdot \frac{x^{m-1}}{(m-1)!} \cdot e^{\lambda x}.$$

This result will prove to be very important in the following computation. It turns out, and this will be shown shortly, that there is a simple general solution to the differential equation $DZ = 0$, given that the differential operator D is of the form described in equation (4.3). Therefore, the inverse of $D\delta$ can always be written explicitly, and as claimed will be an element of $\mathcal{D}'_+$. To see this and to work out the explicit form, consider the polynomial with the same form as the differential operator.

$$\begin{aligned} P(z) &\equiv \sum_{k=0}^{m} a_k \cdot z^k \text{ with } a_m = 1 \\ &= \prod_j (z - z_j)^{k_j}, \end{aligned}$$

where z_j are the roots of $P(z)$ with multiplicity k_j. Then note that

$$D\delta = \circledast_j \; (\delta' - z_j)^{k_j},$$

where '$\circledast$' represents the repeated convolution product. By Theorem 4.18, this is an element of $\mathcal{D}'_+$. Now note that for any two distributions S and T, $S * T = 0 \Rightarrow S = 0$ or $T = 0$. This means that in fact, $\mathcal{D}'_+$ is a quotient field, and while the exact meaning of that is beyond the scope of this report, it yields a property that will be very useful. Namely, that any inverse S^{-1} can be written as $\frac{1}{S}$, i.e. there is an inverse operation to " $*$ " in $\mathcal{D}'_+$. The interested reader can learn more about this field and its properties in [4] (Paragraph 6.7).

This will be used in order to apply partial fraction decomposition to $P(z)$, the usefulness of which will become apparent immediately afterwards. Hence by partial fraction decomposition,

$$\frac{1}{P(z)} = \frac{1}{\prod_j (z - z_j)^{k_j}} = \sum_j \sum_{i=0}^{k_j} \frac{c_{j,i}}{(z - z_j)^i} \text{ for some } (c_{j,i}) \in \mathbb{C}.$$

Recall that by Example 4.19, for $D = (\frac{d}{dx} - \lambda)^m$ it was computed that $(\delta' - \lambda \cdot \delta)^{m^{-1}} = H(x) \cdot \frac{x^{m-1}}{(m-1)!} \cdot e^{\lambda x}$. Note that in general, the differential operator is the convolution product of terms in the form of that in Example 4.19. And by this property of $\mathcal{D}'_+$ being a quotient field, one can thus express the inverse of $D\delta$ by partial fraction decomposition, with the exact same coefficients as the decomposition of $\frac{1}{P(z)}$, since that is the exact same weighted sum. The only difference is the terms that it sums over, but for each individual term, the inverse was already computed in Example 4.19. Combining this gives

$$(D\delta)^{-1} = H(x) \cdot \sum_j \left(\sum_{i=0}^{k_j} c_{j,i} \cdot \frac{x^{i-1}}{(i-1)!} \right) \cdot e^{z_j x}. \tag{4.4}$$

In conclusion, for every differential operator D, there exists an explicitly computable inverse of $D\delta$. As a result thereof, the following corollary is very simple to prove.

Corollary 4.20.

1. Every differential operator D has a unique elementary solution $E \in \mathcal{D}'_+$ of $DE = \delta$, given by $E = (D\delta)^{-1}$.

2. Given a differential operator D, for every $T \in \mathcal{D}'_+$ there exists a unique solution $S \in \mathcal{D}'_+$ of $DS = T$, given by $S = (D\delta)^{-1} * T$.

Proof.

1. Note that $DE = \delta \Leftrightarrow D\delta * E = \delta \Leftrightarrow E = (D\delta)^{-1} * \delta = (D\delta)^{-1}$.

2. Note that $DS = T \Leftrightarrow D\delta * S = T \Leftrightarrow S = (D\delta)^{-1} * T$.

By Theorem 4.16 both solutions are unique, and by Theorem 4.18, $E, S \in \mathcal{D}'_+$. □

For this reason, $\mathcal{D}'_+$ is a very useful space to work in, as then every explicit differential equation can be solved using convolutions, which are guaranteed to have a unique solution in $\mathcal{D}'_+$. As mentioned before, everything can be done in $\mathcal{D}'_-$ as well, the only change is having to substitute $H(x)$ by $-H(-x)$ everywhere.

The procedure of finding $(D\delta)^{-1}$ may look very complex, due to the inconvenient index notation of multiple roots and partial fraction decomposition, but it is really not. Mostly because those two steps are the only ones needed in order to find $(D\delta)^{-1}$, apart from having to fill the coefficients into equation (4.4), that is. To illustrate this, let us treat a simple example, namely finding the elementary solution of a differential operator that one may have encountered before.

Example 4.21. Let $D = \frac{d^2}{dx^2} + \omega^2$, for some fixed $\omega \in \mathbb{R}$. The solution to $y''(x) = -\omega^2 \cdot y(x)$ is exactly the elementary solution of this differential operator, i.e. $y \in \mathcal{D}'_+$ such that $Dy = 0$.

To find it, let us use the process described before. Set $P(z) \equiv z^2 + \omega^2$, then its roots are $\pm\imath\omega$, both with multiplicity 1. Then by partial fraction decomposition,

$$\frac{1}{P(z)} = \frac{1}{z^2 + \omega^2} = \frac{1}{z - \imath\omega} \cdot \frac{1}{z + \imath\omega} = \frac{1}{2\imath\omega} \cdot \left(\frac{1}{z - \imath\omega} - \frac{1}{z + \imath\omega} \right).$$

Hence the partial fraction decomposition has coefficients $\frac{\pm 1}{2 \cdot \imath\omega}$ for the roots $\pm\imath\omega$ respectively. So

$$(D\delta)^{-1} = H(x) \cdot \frac{1}{2\imath\omega} \cdot (e^{\imath\omega x} - e^{-\imath\omega x}) = H(x) \cdot \frac{\sin(\omega x)}{\omega}.$$

5. Fourier Transform

5.1. Fourier Transform of Functions

The first topic that can be seen as an application of distribution theory, is the Fourier Transform. However, except of showing how distribution theory can be used in this field, this chapter will also be used as a stepping stone towards applying distributions in more advanced applications. The experienced reader may be well aware that in many difficult differential equations, performing a Fourier transform in order to achieve an easier expression is often the difference between being able to solve this differential equation or not. First, this chapter will focus on exploring how the Fourier transform is exactly defined on functions, and extend this theory to develop useful theorems and properties, that are often used when working with differential equations. In the second part of this chapter, the Fourier transform and its properties can then seamlessly be extended to act upon distributions. This will be the final preparation before concrete advanced applications of distributions can be explored, while keeping in mind that the second part of this chapter can also be regarded as an application of distribution theory itself.

As a final note, in the entirety of this chapter, computations will be done in the space $\mathbb{R}$. However, everything that will be treated in this chapter can trivially be extended to $\mathbb{R}^n$, but the one-dimensional case might be more intuitive at first, mostly due to notation.

Now, let us first recall the definition of the Fourier transform of functions, a few of its more trivial properties, and treat some examples that will be recurring throughout this chapter.

Definition 5.1. Let $f \in L^1(\mathbb{R})$, then its Fourier transform is defined as

$$\widehat{f}(y) \equiv \int_{-\infty}^{\infty} f(x)\, e^{-2\pi \imath xy}\, dx \text{ for } y \in \mathbb{R}.$$

This is also denoted $\mathcal{F}f$, since the Fourier transform can be seen as a map $\mathcal{F} : L^1(\mathbb{R}) \to \mathcal{E}^0(\mathbb{R})$. Let us now list some properties, the proofs of which can be worked out fairly easily.

Property 5.1. For all $f, g \in L^1(\mathbb{R})$ and for all $a \in \mathbb{R}$:

1. $\mathcal{F} : L^1(\mathbb{R}) \to \mathcal{E}^0(\mathbb{R})$ is linear, but not surjective.
2. $|\mathcal{F}f(y)| \leq \|f\|_1 \; \forall y \in \mathbb{R}$.
3. $\mathcal{F}(f * g) = \mathcal{F}f \cdot \mathcal{F}g$.
4. Let $\tilde{f}(x) \equiv \overline{f(-x)}$, then $\mathcal{F}\tilde{f}(y) = \overline{\mathcal{F}f(y)}$.

A few important and recurring examples of Fourier transformations are given below.

Example 5.2.

1. Let $f_a(x) = e^{-a|x|}$ for $a > 0$, then $\widehat{f_a}(y) = \frac{2a}{a^2+4\pi^2y^2}$.
2. Let $f_a(x) = e^{-ax^2}$ for $a > 0$, then $\widehat{f_a}(y) = \sqrt{\frac{\pi}{a}} \cdot e^{-\frac{\pi^2y^2}{a}}$.
3. $\mathcal{F}(\tau_{(a)}f)(y) = e^{-2\pi\imath ay} \cdot \widehat{f}(y)$ and $\mathcal{F}(e^{2\pi\imath ax} \cdot f)(y) = \tau_{(a)}\widehat{f}(y)$.

Having explored these basic properties of the Fourier transform, it is now possible to prove some of the more complex and important theorems regarding the Fourier transform, all of which will be of great help when applying the Fourier transform on distributions. First and foremost, there is the inverse Fourier transform, the form of which looks very intuitive when looking at Definition 5.1.

Theorem 5.3. Let $f \in L^1(\mathbb{R})$, and assume that at some fixed $x \in \mathbb{R}$,

$$f(x^+) \equiv \lim_{t \downarrow x} f(t) \text{ and } f(x^-) \equiv \lim_{t \uparrow x} f(t) \text{ exist.}$$

Then for all $x \in \mathbb{R}$ for which this is the case,

$$\lim_{a \downarrow 0} \int_{-\infty}^{\infty} \widehat{f}(y)\ e^{2\pi\imath xy}\ e^{-4\pi^2ay^2}\ dy = \frac{f(x^+) + f(x^-)}{2}.$$

In particular, if f is continuous, then

$$\int_{-\infty}^{\infty} \widehat{f}(y)\ e^{2\pi\imath xy}\ dy = f(x)\ \forall x \in \mathbb{R}.$$

The proof of this is surprisingly technical in the case that f is not continuous, and the reader will be referred to [4] (Theorem 7.1) for it.

From the theorem itself though, there are a few things that can immediately be concluded. First of all, in the case that f is continuous, note that the last equation simply states that $\bar{\mathcal{F}}\widehat{f} = f$. Hence the complex conjugate of the Fourier transform is the inverse Fourier transform. In this case, it follows in particular that

$$f(0) = \int_{-\infty}^{\infty} \widehat{f}(y)\ dy. \tag{5.1}$$

This equation can be generalized in order to hold for even more functions, forming a corollary that will be necessary later this chapter.

Corollary 5.4. Let $f \in L^1(\mathbb{R})$ be nonnegative. Then if $f(x)$ is continuous at $x = 0$, it follows that $\widehat{f} \in L^1(\mathbb{R})$, and equation 5.1 holds. Moreover, this still holds when replacing $L^1(\mathbb{R})$ by $L^2(\mathbb{R})$ everywhere.

The proof of the first part is very extensive without giving additional insight, and therefore the reader is redirected to [6] (Theorem 2.30 and Theorem 2.31) for it.

Using this last corollary and the useful observation that $\mathcal{F}^{-1} = \bar{\mathcal{F}}$, it is actually possible to show that the Fourier transform preserves the norm of certain functions. This property can then be generalized into a broader equation, yielding Plancherel's theorem, given below.

Theorem 5.5 (Plancherel)**.** Let $f, g \in L^1(\mathbb{R})$ and assume $f, g \in L^2(\mathbb{R})$. Then

$$\int_{-\infty}^{\infty} \widehat{f}(y)\ \overline{\widehat{g}(y)}\ dy = \int_{-\infty}^{\infty} f(x)\ \overline{g(x)}\ dx.$$

Moreover, $\widehat{f} \in L^2(\mathbb{R})$ and $\|\widehat{f}\|_2 = \|f\|_2$.

Proof. Recall that $\tilde{g}(x) \equiv \overline{g(-x)}$, then by Property 5.1 (items 3. and 4.), $\widehat{f * \tilde{g}} = \widehat{f} \cdot \widehat{\tilde{g}} = \widehat{f} \cdot \overline{\widehat{g}}$. Then applying the inverse Fourier transform gives $f * \tilde{g} = \bar{\mathcal{F}}(\widehat{f} \cdot \overline{\widehat{g}})$. Hence by equation (5.1),

$$\begin{aligned} f * \tilde{g}(0) = \bar{\mathcal{F}}(\widehat{f} \cdot \overline{\widehat{g}(0)}) &= \bar{\mathcal{F}}\left(\widehat{f}(y) \cdot \int_{-\infty}^{\infty} \overline{g(x)}\ dx\right) \\ &= \int_{-\infty}^{\infty} \widehat{f}(y) \int_{-\infty}^{\infty} \overline{g(x)}\ dx\ e^{2\pi\imath xy}\ dy \\ &= \int_{-\infty}^{\infty} \widehat{f}(y) \int_{-\infty}^{\infty} \overline{g(x)}\ e^{2\pi\imath xy}\ dx\ dy \\ &= \int_{-\infty}^{\infty} \widehat{f}(y)\ \overline{\widehat{g}(y)}\ dy. \end{aligned}$$

Where the third line was deduced by using Fubini's theorem. On the other hand, by the definition of convolution for functions,

$$f * \tilde{g}(0) = \int_{-\infty}^{\infty} f(x)\ \overline{g(x)}\ dx,$$

which proves the desired equality. In this equality, substitute $g = f$, then

$$\begin{aligned} \int_{-\infty}^{\infty} \widehat{f}(y)\ \overline{\widehat{f}(y)}\ dy &= \int_{-\infty}^{\infty} f(x)\ \overline{f(x)}\ dx \\ \Rightarrow \int_{-\infty}^{\infty} \left|\widehat{f}(y)\right|^2\ dy &= \int_{-\infty}^{\infty} |f(x)|^2\ dx. \end{aligned}$$

Hence, $\|\widehat{f}\|_2 = \|f\|_2$, and thus, $\widehat{f} \in L^2(\mathbb{R})$. □

This is a central theorem in Fourier analysis. It tells us that the Fourier transform, if restricted to $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$, can be extended to an isometry from $L^2(\mathbb{R})$ to $L^2(\mathbb{R})$. This means that $\|f\|_2 = \|\mathcal{F}f\|_2$, as proven in the last line of the proof. Another consequence of this is that $f = 0 \Leftrightarrow \widehat{f} = 0$ for all $f \in L^2(\mathbb{R})$.

As a final topic in this section, let us discover a few surprising, but incredibly useful, identities of Fourier transforms regarding differentiation. That is, what it means to take the Fourier transform of a derivative and to take the derivative of a Fourier transform, how these are related, and if there exist simple forms of writing and interpreting these. Let us start with the first aforementioned case.

Theorem 5.6. If $f \in \mathcal{E}^1(\mathbb{R})$ is such that $f' \in L^1(\mathbb{R})$, then $\widehat{f'}(y) = 2\pi\imath y \cdot \widehat{f}(y)$.

Proof. For $y \neq 0$, by partial integration on $[-a, a]$,

$$\int_{-a}^{a} f(x)\ e^{-2\pi\imath xy}\ dx = \left[\frac{e^{-2\pi\imath xy}}{-2\pi\imath y} \cdot f(x)\right]_{-a}^{a} + \frac{1}{2\pi\imath y} \cdot \int_{-a}^{a} f'(x)\ e^{-2\pi\imath xy}\ dx. \tag{5.2}$$

Since $f' \in L^1(\mathbb{R})$,

$$\lim_{|x| \to \infty} f(x) = f(0) + \int_0^{\infty} f'(t)\ dt < \infty.$$

Moreover, $f \in L^1(\mathbb{R})$, and thus

$$\lim_{|x|\to\infty} f(x) = 0, \text{ so } \lim_{a\to\infty} [f(x)]_{-a}^{a} = 0.$$

Hence, taking the limit $a \to \infty$ on both sides of equation 5.2,

$$\int_{-\infty}^{\infty} f(x)\, e^{-2\pi \imath xy}\, dx = \frac{1}{2\pi \imath y} \cdot \int_{-\infty}^{\infty} f'(x)\, e^{-2\pi \imath xy}\, dx.$$

Therefore, $\widehat{f}(y) \cdot 2\pi \imath y = \widehat{f'}(y)$ for $y \neq 0$. By continuity of f, this also holds for $y = 0$. □

It is important to realize the peculiarity of this result; the Fourier transform of a derivative can be written as the Fourier transform multiplied by a simple linear function. This can trivially be extended to higher order derivatives, using induction.

Corollary 5.7. Fix $m \in \mathbb{N}$ and let $f \in \mathcal{E}^m(\mathbb{R})$ be such that $f^{(k)} \in L^1(\mathbb{R})$ for all $0 \leq k \leq m$. Then $\mathcal{F}(f^{(m)}(x)) = (2\pi \imath y)^m \cdot \mathcal{F}f(y)$.

The proof of last theorem may be lengthy, but it hopefully provides insight into why this remarkable expression is justified. Moreover, this can be used to estimate what a similar expression for the derivative of a Fourier transform looks like, and the following theorem will confirm such estimation.

Theorem 5.8. Let $f \in L^1(\mathbb{R})$ and suppose $g(x) \equiv -2\pi \imath x \cdot f(x) \in L^1(\mathbb{R})$. Then $\widehat{f} \in \mathcal{E}^1(\mathbb{R})$ and $(\widehat{f})' = \widehat{g}$.

The proof of this, while important, does not provide any additional insight as to why the derivative of the Fourier transform is of this form. However, as mentioned right before the theorem, the general form of the Fourier transform of a derivative may give this insight instead. For the proof itself, however, the reader is redirected to Appendix B (Theorem B.1). Finally, note that this last theorem can be generalized to higher order derivatives as well.

Corollary 5.9. Fix $m \in \mathbb{N}$, let $f \in L^1(\mathbb{R})$ and suppose $g_k(x) \equiv (-2\pi \imath x)^k \cdot f(x) \in L^1(\mathbb{R})$ for all $0 \leq k \leq m$. Then $\widehat{f} \in \mathcal{E}^m(\mathbb{R})$ and $(\widehat{f})^{(m)} = \widehat{g_m}$.

Note that this last equation can be written as $(\mathcal{F}f)^{(m)}(y) = (-2\pi \imath y)^m \cdot \mathcal{F}f(y)$. However, this leads us to easily form a relation between the derivative of the Fourier transform and the Fourier transform of the derivative, by the following observation that in the presupposed conditions;

$$(\mathcal{F}f)^{(m)}(y) = (-2\pi \imath y)^m \cdot \mathcal{F}f(y) = (-1)^m \mathcal{F}(f^{(m)}(x)) \tag{5.3}$$

5.2. Tempered Distributions

Now that an extensive recollection of the Fourier transform of functions has been made, and this has been elaborately expanded by several useful theorems, it is almost possible to define the Fourier transform of distributions. In order to provide an educated guess of its definition, let us first consider how the Fourier transform of a regular distribution would behave. It turns out that this gives a satisfying result, as will be seen below.

Consider $f \in L^1(\mathbb{R})$, then clearly for all $\varphi \in \mathcal{D}(\mathbb{R})$,

$$\begin{aligned}\langle \widehat{f}, \varphi \rangle &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x)\ e^{-2\pi \imath xy}\ dx\ \varphi(y)\ dy \\ &= \int_{-\infty}^{\infty} f(x) \int_{-\infty}^{\infty} \varphi(y)\ e^{-2\pi \imath xy}\ dy\ dx \\ &= \langle f, \widehat{\varphi} \rangle\end{aligned}$$

Hence for any $T_f \in \mathcal{D}'(\mathbb{R})$, $\langle \widehat{T_f}, \varphi \rangle = \langle T_f, \widehat{\varphi} \rangle$ for $\varphi \in \mathcal{D}(\mathbb{R})$. Thus, a reasonable definition of the Fourier transform of a distribution $T \in \mathcal{D}'(\mathbb{R})$ would be that $\langle \widehat{T}, \varphi \rangle \equiv \langle T, \widehat{\varphi} \rangle$ for $\varphi \in \mathcal{D}(\mathbb{R})$. While this is mostly correct, there is one technicality that turns out to be a problem. Namely, is it even guaranteed that $\widehat{\varphi} \in \mathcal{D}(\mathbb{R})$? To find out, note that for $\operatorname{supp} \varphi \subset [-a, a]$,

$$\widehat{\varphi}(y) = \int_{-a}^{a} \varphi(x)\ e^{-2\pi \imath xy}\ dx = \sum_{k=0}^{\infty} \frac{(-2\pi \imath x)^k}{k!} \cdot \int_{-a}^{a} \varphi(x)\ x^k\ dx.$$

Assume that $\widehat{\varphi} \in \mathcal{D}(\mathbb{R})$, then $(\widehat{\varphi})^{(m)}(x)$ is continuous for all $m \in \mathbb{N}$, and thus in order for the integral to exist, $\varphi(x)$ must be 0. Usually this is not the case, hence $\widehat{\varphi}$ is usually not in $\mathcal{D}(\mathbb{R})$. So instead, let us define a new space, denoted $\mathcal{S}(\mathbb{R})$, such that $\mathcal{F} : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$, defined below.

Definition 5.10. The Schwartz space on $\mathbb{R}$, denoted $\mathcal{S}(\mathbb{R})$, consists of all $\varphi \in \mathcal{E}(\mathbb{R})$ such that $x^l \cdot \varphi^{(k)}(x)$ is bounded on $\mathbb{R}$ for all $k, l \in \mathbb{N}$.

This definition can trivially be extended to act on any space other than $\mathbb{R}$. Also, note that $\mathcal{D} \subset \mathcal{S} \subset \mathcal{E}$. Although the definition of this space is fairly straightforward, let us quickly observe that the function $\varphi(x) = e^{-ax^2}$ is in $\mathcal{S}(\mathbb{R})$ for all $a > 0$. Now, there are a few trivial, but important properties of this space to realize, listed below.

Property 5.2.

1. $\mathcal{S}$ is a vector space.
2. If $\varphi \in \mathcal{S}$, then $x^l \cdot \varphi^{(k)}(x) \in \mathcal{S}$ and $(x^l \cdot \varphi(x))^{(k)} \in \mathcal{S}$ for any $k, l \in \mathbb{N}$.
3. $\mathcal{S} \subset L^p$ for all $1 \leq p \leq \infty$.
4. $\mathcal{F}(\mathcal{S}) = \mathcal{S}$.

And note that all of these hold for any space that $\mathcal{S}$ may act on. The final property may not look clear cut at first sight, however note that by Corollaries 5.7 and 5.9, Fourier transforms of derivatives are of the form presented in the second line of this Property. This is one of the most important ingredients for this proof, in order to fill in the (very technical) gaps, the reader is referred to [3] (Lemma 5.12).

Now, as has been done with all other spaces defined before, and as will be necessary for the next theorem, let us define convergence on the spaces $\mathcal{S}$ and $\mathcal{S}'$, again acting on any space.

Definition 5.11. A sequence $(\varphi_j) \in \mathcal{S}$ converges to 0 if $\sup |x^l \varphi_j^{(k)}(x)| \xrightarrow{\mathbb{C}} 0$ for all $k, l \in \mathbb{N}$. This is denoted $\varphi_j \xrightarrow{\mathcal{S}} 0$. Moreover, a sequence $(T_j) \in \mathcal{S}'$ converges to 0 if $\langle T_j, \varphi \rangle \xrightarrow{\mathbb{C}} 0$ for all $\varphi \in \mathcal{S}$. This is denoted $T_j \xrightarrow{\mathcal{S}'} 0$.

Note that both of these notions can be extended trivially, by noting that $\varphi_j \xrightarrow{\mathcal{S}'} \psi$ is equivalent to $(\varphi_j - \psi) \xrightarrow{\mathcal{S}'} 0$, the latter of which is defined. The exact same holds for the second notion. Also note that whenever $T_j \xrightarrow{\mathcal{D}'} 0$, it follows that $T_j \xrightarrow{\mathcal{S}'} 0$ for any $T_j \in \mathcal{S}'$. This shows that the previous definition is indeed consistent with Definition 3.23.

Let us return to $\mathbb{R}$ for now, in order to keep insights intuitive and the proofs clear. Recall that $\mathcal{F} : L^1(\mathbb{R}) \to \mathcal{E}^0(\mathbb{R})$ is a non-surjective isometry. Also note that by last proposition, $\mathcal{F} : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$ is linear and bijective. Using last definition, it is possible to prove continuity of this Fourier transform, the consequence of which is stated in the following theorem.

Theorem 5.12. $\mathcal{F} : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$ is an isomorphism.

The proof of the necessary continuity however, is very technical and the reader will be referred to [4] (Theorem 7.11) instead.

Recalling that $\mathcal{D} \subset \mathcal{S}$, and thus $\mathcal{S}' \subset \mathcal{D}'$, let us classify the distributions that can be extended to a continuous linear form on $\mathcal{S}$.

Definition 5.13. $T \in \mathcal{D}'(\mathbb{R})$ is called a tempered distribution if it can be extended to a continuous linear form $L \in \mathcal{S}'(\mathbb{R})$.

Usually it is denoted that in the presupposed case, $T \in \mathcal{S}'(\mathbb{R})$, as the difference between a tempered distribution and an element of $\mathcal{S}'(\mathbb{R})$ is merely one of notation.

Even though the definition of $\mathcal{S}(\mathbb{R})$ is straightforward, the structure of $\mathcal{S}'(\mathbb{R})$ may not be. Hence let us consider two general cases in which the given class of distributions is also tempered.

Example 5.14.

- Every distribution with compact support can be extended to $\mathcal{E}'(\mathbb{R})$ (see Theorem 3.26), then restricted to $\mathcal{S}'(\mathbb{R})$, and is thus a tempered distribution. This reinforces the observation that $\mathcal{D}(\mathbb{R}) \subset \mathcal{S}(\mathbb{R}) \subset \mathcal{E}(\mathbb{R})$, and thus indeed $\mathcal{E}'(\mathbb{R}) \subset \mathcal{S}'(\mathbb{R}) \subset \mathcal{D}'(\mathbb{R})$.

- Fix $p \in \mathbb{N}$ with $1 \leq p \leq \infty$ and let $f \in L^p(\mathbb{R})$. Then for $q \in \mathbb{N}$ such that $\frac{1}{p} + \frac{1}{q} = 1$, T_f is tempered and $\widehat{f} \in L^q(\mathbb{R})$. This is consistent with Property 5.2, which states that $\mathcal{S}(\mathbb{R}) \subset L^q(\mathbb{R})$. The computation of this, however, falls outside the scope of this report, and thus the reader is referred to [7] (Paragraph 1.3) for this.

Finally, it turns out that there is a condition for a distribution being tempered, similar to that of Proposition 3.5 for identifying distributions and Proposition 3.21 for checking whether a distribution has compact support, and its proof is similar too.

Proposition 5.15. A distribution T is tempered if and only if there exist constants $C > 0$ and $m \in \mathbb{N}$, such that

$$|\langle T, \varphi \rangle| \leq C \cdot \sum_{k,l \leq m} \sup |x^l \cdot \varphi^{(k)}(x)| \;\; \forall \varphi \in \mathcal{D}(\mathbb{R})$$

5.3. Fourier Transform of Distributions

At long last, it is possible to define the Fourier transform of a distribution in a mathematically solid way, a hint of which has already been given at the start of section 5.2. The major difference, of course, is the fact that the Fourier transform is taken of a tempered distribution.

Definition 5.16. Let $T \in \mathcal{S}'(\mathbb{R})$, then its Fourier transform $\widehat{T}$ is defined by $\langle \widehat{T}, \varphi \rangle \equiv \langle T, \widehat{\varphi} \rangle$ for $\varphi \in \mathcal{S}(\mathbb{R})$.

Note that since $\widehat{\varphi} \in \mathcal{S}(\mathbb{R})$, $\widehat{T} \in \mathcal{S}'(\mathbb{R})$. Also the inverse Fourier transform $\bar{\mathcal{F}}$ is defined similarly and analogously to Theorem 5.12, $\mathcal{F} : \mathcal{S}'(\mathbb{R}) \to \mathcal{S}'(\mathbb{R})$ is an isomorphism.

Let us now show that this definition is consistent with the observation at the start of section 5.2, regarding the Fourier transform of a regular distribution, and even extend this observation.

Property 5.3. If $f \in L^1(\mathbb{R})$ or $f \in L^2(\mathbb{R})$, then $\widehat{T}_f = T_{\widehat{f}}$.

Proof. In the case that $f \in L^1(\mathbb{R})$, it follows from Fubini's theorem that

$$\begin{aligned}
\langle \widehat{T}_f, \varphi \rangle = \langle f, \widehat{\varphi} \rangle &= \int_{-\infty}^{\infty} f(y) \int_{-\infty}^{\infty} \varphi(x)\, e^{-2\pi\imath xy}\, dx\, dy \\
&= \int_{-\infty}^{\infty} \varphi(x) \int_{-\infty}^{\infty} f(y)\, e^{-2\pi\imath xy}\, dy\, dx \\
&= \langle \widehat{f}, \varphi \rangle = \langle T_{\widehat{f}}, \varphi \rangle.
\end{aligned}$$

The case in which $f \in L^2(\mathbb{R})$ now follows from Plancherel's theorem. □

In order to get a sense of how some Fourier transforms of distributions are explicitly defined, let us compute the Fourier transforms of the Dirac delta, and a few variations thereof.

Example 5.17. First, note that

$$\begin{aligned}
\langle \mathcal{F}\delta, \varphi \rangle = \langle \delta, \mathcal{F}\varphi \rangle &= \widehat{\varphi}(0) \\
&= \int_{-\infty}^{\infty} \varphi(x)\, e^{-2\pi\imath xy}\, dx\, \Big|_{y=0} \\
&= \int_{-\infty}^{\infty} \varphi(x)\, dx = \langle 1, \varphi \rangle \text{ for } \varphi \in \mathcal{S}(\mathbb{R}).
\end{aligned}$$

Hence it follows that $\mathcal{F}\delta = T_1$. Clearly, then $\bar{\mathcal{F}}T_1 = \delta$, but

$$\begin{aligned}
\langle \bar{\mathcal{F}}T_1, \varphi \rangle = \langle \bar{\mathcal{F}}1, \varphi \rangle &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{2\pi\imath xy}\, dy\, \varphi(x)\, dx \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-2\pi\imath xy}\, dy\, \varphi(x)\, dx \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \varphi(x)\, e^{-2\pi\imath xy}\, dx\, dy \\
&= \langle 1, \varphi \rangle = \langle \mathcal{F}T_1, \varphi \rangle \text{ for } \varphi \in \mathcal{S}(\mathbb{R}).
\end{aligned}$$

Thus, $\delta = \mathcal{F}T_1$. Generalizing this example gives $\mathcal{F}\delta_{(a)} = T_{e^{2\pi\imath ay}}$. Now, consider

$$\begin{aligned}\langle \mathcal{F}\delta', \varphi\rangle &= \langle \delta', \mathcal{F}\varphi\rangle = -(\mathcal{F}\varphi)'(0)\\ &= \frac{d}{dy}\int_{-\infty}^{\infty} -\varphi(x)\ e^{-2\pi\imath xy}\ dx\ \Big|_{y=0}\\ &= \int_{-\infty}^{\infty} 2\pi\imath x\ \varphi(x)\ dx\\ &= \langle 2\pi\imath x, \varphi\rangle = \langle T_{2\pi\imath x}, \varphi\rangle \text{ for } \varphi \in \mathcal{S}(\mathbb{R}).\end{aligned}$$

So $\mathcal{F}\delta' = T_{2\pi\imath x}$, which since $\widehat{\delta} = 1$ is very analogous to $\widehat{f}'(x) = 2\pi\imath x \cdot \widehat{f}(x)$. This too can be generalized, to $\mathcal{F}\delta^{(m)} = T_{(2\pi\imath x)^m}$.

Following this example, it is not hard to check that Corollaries 5.7 and 5.9 can be extended to distributions, and that equation (5.3) can be extended to

$$(\mathcal{F}T)^{(m)} = (-2\pi\imath y)^m \cdot \mathcal{F}T = (-1)^m \mathcal{F}(T^{(m)})$$

It is now clear how differentiability and the Fourier transform are related, the results of which will turn out to be very useful in the upcoming applications. The final field within the theory of Fourier transforms is the following, regarding slowly increasing functions. Now, this may look like a sudden switch, but it will turn out to give incredible results, that dovetail with the rest of this section. Let us define what this notion is, and then explore what it leads to.

Definition 5.18. A function $f \in L^1(\mathbb{R})$ is called slowly increasing if there exist constants $C > 0$ and $k \in \mathbb{N}$, such that $|f(x)| \le C \cdot |x|^k$ for $|x|$ sufficiently large.

Note that for any slowly increasing function f, from Property 5.15 it follows that $T_f \in \mathcal{S}'(\mathbb{R})$. There is however a remarkable theorem that shows the opposite as well.

Theorem 5.19. Let $T \in \mathcal{S}'(\mathbb{R})$, then there exist some $f \in \mathcal{E}$ which is slowly increasing, and $m \in \mathbb{N}$ such that $T = T_{f^{(m)}}$.

The proof of this theorem is very technical, using various advanced theorems of the field of analysis, therefore the reader is referred to [4] (Theorem 7.14) for this.

Nevertheless, this is a very important result, namely that any tempered distribution is equal to a regular distribution, defined by some derivative of a slowly increasing function. Unfortunately, by the absence of its proof, this may not be very insightful. However, it leads us to hypothesize an even more important and general theorem, the proof of which will be laid out in its full extent.

Theorem 5.20. For any $T \in \mathcal{E}'(\mathbb{R})$, the slowly increasing function $f(x) \equiv \langle T, e^{-2\pi\imath xy}\rangle$ defines the Fourier transform of T, i.e. $\widehat{T} = T_f$.

As indicated, this proof is very important to understand, due to the impact of the result. Because to its length however, the proof is displayed in Appendix B (Theorem B.2), which the reader is advised to go over.

This crucial theorem leads to the following useful corollaries, regarding the Fourier transform of a convolution product.

Corollary 5.21. Let $f, g \in L^2(\mathbb{R})$, then $\widehat{T_f * T_g} = T_{\widehat{f} \cdot \widehat{g}}$.

Proof. First recall that for any $h \in L^2(\mathbb{R})$, by Example 5.14 $T_h \in \mathcal{S}'(\mathbb{R})$, and by Property 5.3, $\widehat{T}_h = T_{\widehat{h}}$. Then by the previous theorem, $\widehat{T_f * T_g} = T_h$, where

$$h = \langle f * g, e^{-2\pi \imath x y} \rangle = \widehat{f * g} = \widehat{f} \cdot \widehat{g},$$

and the last equality follows from Property 5.1. Thus, $\widehat{T_f * T_g} = T_{\widehat{f} \cdot \widehat{g}}$. □

This can be generalized to a similar equality for distributions with compact support, relying further on Theorem 5.20. The statement of the theorem itself is a very sloppy version that may be misleading. Therefore, the proof will contain a more mathematically correct version of this statement.

Corollary 5.22. Let $S, T \in \mathcal{E}'(\mathbb{R})$, then $\widehat{S * T} = \widehat{S} \cdot \widehat{T}$.

Proof. First, note that this notation is very sloppy, since the right-hand side is not well-defined, but the actual identity and last corollary may give insight as to why this notation is usually preferred. In order to keep track of every variable, let $U, V \subset \mathbb{R}$ and $Y \equiv U \times V$.

Then for $S \in \mathcal{E}'(U)$ and $T \in \mathcal{E}'(V)$, it follows from Definition 4.6 that $(S * T) \in \mathcal{E}'(Y)$. Moreover, by Theorem 5.20, $\widehat{S * T} = T_f$, where

$$\begin{aligned} f &= \langle S * T, e^{-2\pi \imath x y} \rangle = \langle S \otimes T, e^{-2\pi \imath x(u+v)} \rangle \\ &= \langle S, e^{-2\pi \imath x u} \cdot \langle T, e^{-2\pi \imath x v} \rangle \rangle \\ &= \langle S, e^{-2\pi \imath x u} \rangle \cdot \langle T, e^{-2\pi \imath x v} \rangle \\ &= g \cdot h, \text{ where } \widehat{S} = T_g \text{ and } \widehat{T} = T_h. \end{aligned}$$

The first line follows from Definition 4.6 and the last line from Theorem 5.20. Hence it follows that $T_f = T_{g \cdot h}$, where $T_f = \widehat{S * T}$, $T_g = \widehat{S}$ and $T_h = \widehat{T}$. This is usually denoted $\widehat{S * T} = \widehat{S} \cdot \widehat{T}$, even though this is not well-defined. □

As a final note, this last corollary can be extended to the case in which either distribution (or both) is in $\mathcal{S}'(\mathbb{R})$. Finally, recall that anything in this last chapter also holds for $\mathcal{S}'(X)$, for any subset $X \subset \mathbb{R}^n$, usually $\mathbb{R}^n$.

6. Applications

6.1. Fourier Transform

As stated in last chapter, the domain of the Fourier transform can trivially be extended from $\mathbb{R}$ to $\mathbb{R}^n$, although especially $n = 1$ and $n = 3$ will be used in the applications below. Let $x = (x_1, \dots, x_n), y = (y_1, \dots, y_n) \in \mathbb{R}^n$, then $dx \equiv dx_1\ dx_2\ \dots\ dx_n$ and furthermore, for $f(x) \in L^1$,

$$\langle x, y\rangle \equiv \sum_{i=1}^{n} x_i \cdot y_i \text{ and } \widehat{f}(y) \equiv \int_{\mathbb{R}^n} f(x)\ e^{-2\pi\imath\langle x,y\rangle}\ dx \text{ for } y \in \mathbb{R}^n.$$

As a final note, for $r \equiv \sqrt{x_1^2 + \dots + x_n^2}$, the Fourier transform is as in $\mathbb{R}$. For now, let us also introduce a final concept below.

Definition 6.1. A function $f : \mathbb{R}^n \to \mathbb{C}$ is called radial if $f(x_1, \dots, x_n) = \varphi(r)$ for some function $\varphi : \mathbb{R} \to \mathbb{C}$.

In this case, $f(x) \in L^1$ if and only if $r^{n-1} \cdot \varphi(r) \in L^1(\langle 0, \infty\rangle)$. Moreover, $\widehat{f}(y) = \langle f(x), e^{2\pi\imath\langle x,y\rangle}\rangle = \langle \varphi(r), e^{2\pi\imath r|y|}\rangle$, a property that will be used in section 6.5.

Now, let us treat a very simple example to get familiarized with the use of Fourier transforms in solving differential equations, before moving on to more extensive and complex applications.

Example 6.2. Consider the functions $u(x) \in \mathcal{E}^2(\mathbb{R})$ and $f(x) \in L^1(\mathbb{R})$, satisfying

$$-u''(x) + u(x) = f(x).$$

Taking the Fourier transform of both sides gives

$$\begin{aligned} 4\pi^2 y^2 \cdot \widehat{u}(y) + \widehat{u}(y) &= \widehat{f}(y) \\ \Rightarrow \widehat{u}(y) &= \frac{1}{1 + 4\pi^2 y^2} \cdot \widehat{f}(y), \end{aligned}$$

recalling that $\mathcal{F}(u''(x)) = -4\pi^2 y^2 \cdot \mathcal{F}u(y)$, by equation (5.3). Also, from the first line in Example 5.2, it follows that $\bar{\mathcal{F}}\left(\frac{1}{1+4\pi^2 y^2}\right) = \frac{1}{2}\ e^{-|x|}$. Thus,

$$u(x) = \frac{1}{2}\ e^{-|x|} * f(x) = \frac{1}{2} \cdot \int_{-\infty}^{\infty} e^{-|x-y|}\ f(y)\ dy.$$

Adding to this the homogeneous solution $u_{hom}(x) = A \cdot e^x + B \cdot e^{-x}$ gives the general solution

$$u(x) = \frac{1}{2} \cdot \int_{-\infty}^{\infty} e^{-|x-y|}\ f(y)\ dy + A \cdot e^x + B \cdot e^{-x},$$

where $A, B \in \mathbb{C}$ are to be determined by initial conditions for $u(x)$ and $u'(x)$. Moreover, note that $T_f \in \mathcal{S}'(\mathbb{R})$ and that $T_u \in \mathcal{S}'(\mathbb{R})$ if and only if $A = B = 0$, since $e^{\pm x} \notin L^1(\mathbb{R})$.

The central concept of this example, transforming the differential equation to a simpler one using the Fourier transform, is the primary use of the Fourier transform. Not only does this enable us to give an easy solution using convolution, it is sometimes even the only way to give the general solution of a differential equation.

6.2. Particle Motion in Air

Consider a particle M, with mass $m > 0$ and without any dimensions, embedded in the (x, y)-plane. Let its position at a certain time be given by $r(t) \in \mathcal{E}^2(\mathbb{R})$ and its velocity by $v(t) \equiv \frac{dr}{dt}(t)$. Clearly, the total force on the particle must equal $m \cdot \frac{d^2r}{dt^2}$ for $t \geq 0$ by Newton's second law. This introduces a technicality however, namely that it is convenient to redefine the aforementioned vectors for $t \geq 0$ only. Hence, let

$$\boldsymbol{r}(t) \equiv H(t) \cdot r(t) = \begin{cases} r(t) \text{ for } t \geq 0 \\ 0 \text{ for } t < 0 \end{cases} \in \mathcal{D}'_+.$$

It is then very important to note that for differentiation, the product rule applies, i.e.

$$\begin{aligned} \frac{d\boldsymbol{r}}{dt} &= H \cdot \frac{dr}{dt} + r_0 \cdot \delta \\ \frac{d^2\boldsymbol{r}}{dt^2} &= H \cdot \frac{d^2r}{dt^2} + v_0 \cdot \delta + r_0 \cdot \delta'. \end{aligned} \tag{6.1}$$

Recalling that $H' = \delta$ (see Example 3.15 to review how this is applied here), and defining $r_0 \equiv r(0), v_0 \equiv v(0)$. See also Figure 6.1 below for a clear overview.

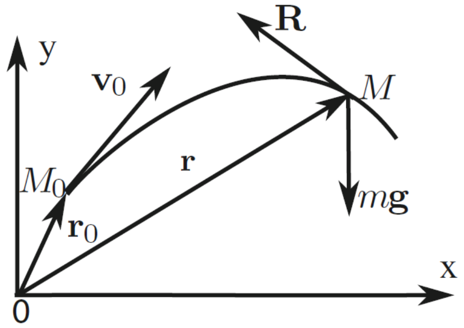

Figure 6.1: Motion of particle M in air (source: [8], Figure 5.2)

In this application, the particle M experiences air resistance, defined by $R(t) \equiv -km \cdot v(t)$ for some $k > 0$. Then, denoting the gravitation constant by g,

$$\begin{aligned} m \cdot \frac{d^2r}{dt^2} &= mg - km \cdot \frac{dr}{dt} \\ \Rightarrow \frac{d^2r}{dt^2} &+ k \cdot \frac{dr}{dt} = g \\ \Rightarrow H \cdot \frac{d^2r}{dt^2} &+ k \cdot H \cdot \frac{dr}{dt} = H \cdot g. \end{aligned}$$

Then by equation (6.1),

$$\frac{d^2\boldsymbol{r}}{dt^2} + k \cdot \frac{d\boldsymbol{r}}{dt} = H \cdot g + (v_0 + kr_0) \cdot \delta + r_0 \cdot \delta'.$$

Let $D = \frac{d^2}{dt^2} + k \cdot \frac{d}{dt}$, then

$$D\delta * \boldsymbol{r} = H \cdot g + (v_0 + kr_0) \cdot \delta + r_0 \cdot \delta'.$$

Using Theorem 4.18, $(D\delta)^{-1} = H \cdot Z$, for $Z(t) \in \mathcal{E}(\mathbb{R})$ satisfying

$$DZ(t) = Z''(t) + k \cdot Z'(t) = 0 \text{ such that } Z(0) = 0, Z'(0) = 1,$$

the solution of which is $Z(t) = \frac{1}{k} \cdot (1 - e^{-kt})$, and thus $E \equiv (D\delta)^{-1} = \frac{H}{k} \cdot (1 - e^{-kt})$. Hence,

$$\begin{aligned} \boldsymbol{r}(t) &= E * (H \cdot g + (v_0 + kr_0) \cdot \delta + r_0 \cdot \delta') \\ &= (E * H) \cdot g + (v_0 + kr_0) \cdot E + r_0 \cdot E'. \end{aligned} \tag{6.2}$$

Then, note that

$$\begin{aligned} E' &= \frac{d}{dt}\left(\frac{H}{k} \cdot (1 - e^{-kt})\right) \\ &= \frac{\delta}{k} \cdot (1 - e^{-kt}) + H \cdot e^{-kt} = H \cdot e^{-kt}, \end{aligned} \tag{6.3}$$

by Example 3.15, as $(1 - e^{-kt})|_{t=0} = 0$. Furthermore,

$$\begin{aligned} E * H &= \left(\frac{H}{k} \cdot (1 - e^{-kt})\right) * H \\ &= \frac{1}{k} \cdot \int_{-\infty}^{\infty} (1 - e^{-k(t-\tau)})\ H(t-\tau)\ H(\tau)\ d\tau \\ &= \frac{1}{k} \cdot \int_{0}^{\infty} (1 - e^{-k(t-\tau)})\ H(t-\tau)\ d\tau. \end{aligned}$$

Substituting $s = t - \tau$ gives

$$\begin{aligned} E * H &= \frac{1}{k} \cdot \int_{-\infty}^{t} (1 - e^{-ks}) \cdot H(s)\ ds \\ &= \frac{1}{k} \cdot \int_{0}^{t} 1 - e^{-ks}\ ds \text{ for } t \geq 0, \end{aligned}$$

and in the case that $t < 0$, $E * H = 0$. Thus,

$$\begin{aligned} E * H &= \frac{H}{k} \cdot \int_{0}^{t} 1 - e^{-ks}\ ds \\ &= \frac{H}{k^2} \cdot (kt - 1 + e^{-kt}) \end{aligned} \tag{6.4}$$

The results of equations (6.3) and (6.4) can be substituted back into equation (6.2), yielding

$$\begin{aligned} \boldsymbol{r}(t) &= (E * H) \cdot g + (v_0 + kr_0) \cdot E + r_0 \cdot E' \\ &= \frac{H \cdot g}{k^2} \cdot (kt - 1 + e^{-kt}) + \frac{H}{k} \cdot (1 - e^{-kt}) \cdot (v_0 + kr_0) + H \cdot e^{-kt} r_0 \\ &= H \cdot \Big(\frac{g}{k^2} \cdot (kt - 1 + e^{-kt}) + \frac{v_0}{k} \cdot (1 - e^{-kt}) + r_0\Big). \end{aligned}$$

And thus,

$$r(t) = \frac{g}{k^2} \cdot (kt - 1 + e^{-kt}) + \frac{v_0}{k} \cdot (1 - e^{-kt}) + r_0 \text{ for } t \geq 0.$$

Note that using convolution, a general solution of the differential equation was easy to formulate. Thereafter, it was simply a matter of computing each term within this solution, in order to express it without using convolutions. As mentioned in chapter 4, this is how convolutions are often used; as a tool to compute the solution easily, but not necessarily appearing in the solution itself. Let us now look at a special case of this differential equation, the solution of which can trivially be derived from the general solution of this application.

Special Case: Trajectory Motion

Consider a special case of this application, in which $r_0 = (0, h)$ and $v_0 = (s, 0)$, for $h, s > 0$. That is, the particle starts at height h along the y-axis and with speed s along the positive x-axis. Then it is possible to describe the motion along both axes, i.e. for $r(t) = (x(t), y(t))$,

$$\begin{aligned} x(t) &= \frac{s}{k} \cdot (1 - e^{-kt}) \\ y(t) &= -\frac{g}{k^2} \cdot (kt - 1 + e^{-kt}) + h, \end{aligned} \tag{6.5}$$

for $t \geq 0$. From the first equation of (6.5) it follows that

$$e^{-kt} = \frac{s - kx}{s} \Rightarrow kt = \log\left(\frac{s}{s - kx}\right).$$

Substituting this into the second line of (6.5), gives

$$y(x) = h - \frac{g}{k^2} \cdot \left(\log\left(\frac{s}{s - kx}\right) - \frac{kx}{s}\right),$$

for $x \in [0, l]$, where $l > 0$ is such that $y(l) = 0$.

6.3. Resonance of a Linear Oscillator

The following application is another one appearing in mechanics, this one regarding the linear oscillator. In particular, it will soon be shown how to derive resonance in such oscillator, and how this leads to an unbounded trajectory. Again, consider a dimensionless particle M of mass m, embedded on the x-axis, and a spring coupling this particle to the origin (see Figure 6.2 below). The position of this particle will be given by $x(t) \in \mathcal{E}^2(\mathbb{R})$ An elastic force pulls M towards the origin, and by Hooke's law, this force is equal to $F = -k \cdot x(t)$ for $k > 0$. Furthermore, an external force acts upon the particle, given by $Q = m \cdot q(t)$ for some $q(t) \in L^1(\mathbb{R})$.

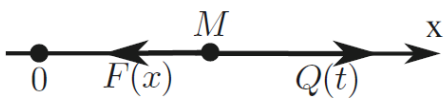

Figure 6.2: Oscillation of the particle M
(source: [8], Figure 5.3)

As in previous example, it is necessary to only consider the positive-time part of each function, i.e. let $\boldsymbol{x}(t) \equiv H(t) \cdot x(t) \in \mathcal{D}'_+$ and $\boldsymbol{q}(t) \equiv H(t) \cdot q(t) \in \mathcal{D}'_+$, then similar to equation (6.1),

$$\begin{aligned} \frac{d\boldsymbol{x}}{dt} &= H \cdot \frac{dx}{dt} + x_0 \cdot \delta \\ \frac{d^2\boldsymbol{x}}{dt^2} &= H \cdot \frac{d^2x}{dt^2} + v_0 \cdot \delta + x_0 \cdot \delta'. \end{aligned} \tag{6.6}$$

As usual, $v(t) \equiv \frac{dx}{dt}(t)$ is the velocity of the particle, and $x_0 \equiv x(0)$, $v_0 \equiv v(0)$ are the initial position and velocity of the particle respectively. Equating all acting forces upon the particle, gives

$$\begin{aligned} m \cdot \frac{d^2x}{dt} + k \cdot x &= m \cdot q \\ \Rightarrow \frac{d^2x}{dt} + \omega^2 \cdot x &= q, \end{aligned}$$

by defining $\omega \equiv \sqrt{\frac{k}{m}} > 0$. Taking only the positive-time part of both sides of these equation, and then applying equation (6.6),

$$\begin{aligned} H \cdot \frac{d^2x}{dt} + \omega^2 \cdot H \cdot x = H \cdot q \\ \Rightarrow \frac{d^2\boldsymbol{x}}{dt} + \omega^2 \cdot \boldsymbol{x} = \boldsymbol{q} + v_0 \cdot \delta + x_0 \cdot \delta' \end{aligned}$$

Consider the differential operator $D = \frac{d^2}{dt^2} + \omega^2$, then by Example 4.21, $E \equiv (D\delta)^{-1} = \frac{H}{\omega} \cdot \sin(\omega t)$, and thus

$$\begin{aligned} \boldsymbol{x}(t) &= E * (\boldsymbol{q} + v_0 \cdot \delta + x_0 \cdot \delta') \\ &= E * \boldsymbol{q} + v_0 \cdot E + x_0 \cdot E' \end{aligned} \tag{6.7}$$

And note that the terms in this equation are given by

$$E' = \frac{\delta}{\omega} \cdot \sin(\omega t) + H \cdot \cos(\omega t) = H \cdot \cos(\omega t),$$

since $\sin(\omega t)|_{t=0} = 0$. Also,

$$\begin{aligned} E * \boldsymbol{q} &= \left(\frac{H}{\omega} \cdot \sin(\omega t)\right) * (H \cdot q) \\ &= \frac{1}{\omega} \cdot \int_{-\infty}^{\infty} \sin(\omega(t-\tau))\ H(t-\tau)\ H(\tau)\ q(\tau)\ d\tau \\ &= \frac{1}{\omega} \cdot \int_{0}^{\infty} \sin(\omega(t-\tau))\ H(t-\tau)\ q(\tau)\ d\tau \\ &= \frac{H}{\omega} \cdot \int_{0}^{t} \sin(\omega(t-s))\ q(s)\ ds \\ &= \frac{H}{\omega} \cdot (\sin(\omega t) * q(t)), \end{aligned}$$

where the fourth line was computed as in section 6.2. Substituting this back into equation (6.7), gives

$$\begin{aligned} \boldsymbol{x}(t) &= E * \boldsymbol{q} + v_0 \cdot E + x_0 \cdot E' \\ &= \frac{H}{\omega} \cdot \int_0^t \sin(\omega(t-s))\ q(s)\ ds + v_0 \cdot \frac{H}{\omega} \cdot \sin(\omega t) + x_0 \cdot H \cdot \cos(\omega t). \end{aligned}$$

Therefore, in the case that $t \geq 0$,

$$x(t) = \frac{1}{\omega} \cdot \int_0^t \sin(\omega(t-s))\ q(s)\ ds + \frac{v_0}{\omega} \cdot \sin(\omega t) + x_0 \cdot \cos(\omega t).$$

Now, contrary to the previous application, the convolution remains present in the final solution. This usually indicates (and is in fact the case here), that this procedure was not only a convenient way to give the solution to this differential equation, but in fact necessary to obtain this first part of that solution. Classically, the homogeneous version of this equation is usually considered, and its solution is indeed the second part of the obtained general solution, which is another indication that this procedure has yielded the correct solution. Now, let us look at two remarkable special cases of this application.

Special Case: Initial External Force

A very simple special case of this application, is that of an initial external pull on the particle. Concretely, let $q(t) = \nu \cdot \delta(t)$, i.e. the particle M is initially pulled away from the origin with speed $\nu > 0$. In this case, for $t \geq 0$,

$$\begin{aligned} x(t) &= \frac{\nu}{\omega} \cdot \int_0^t \sin(\omega(t-s))\ \delta(s)\ ds + \frac{v_0}{\omega} \cdot \sin(\omega t) + x_0 \cdot \cos(\omega t) \\ &= \frac{v_0 + \nu}{\omega} \cdot \sin(\omega t) + x_0 \cdot \cos(\omega t). \end{aligned}$$

This result makes perfect physical sense, as the speed of this initial pull just adds to the initial velocity in the positive-x direction of the particle M.

Special Case: Resonance

A more interesting special case is when the function defining the external force takes the form $q(t) = a \cdot \sin(\omega t)$, for some $a > 0$. In other words, the external force is an oscillation, with the same period as the oscillation that the particle has without any external force. In this setting, for $t \geq 0$,

$$x(t) = \frac{a}{\omega} \cdot \int_0^t \sin(\omega(t-s))\ \sin(\omega s)\ ds + \frac{v_0}{\omega} \cdot \sin(\omega t) + x_0 \cdot \cos(\omega t). \tag{6.8}$$

Working out the integral gives

$$\begin{aligned} \int_0^t \sin(\omega(t-s))\ \sin(\omega s)\ ds &= \frac{1}{2} \cdot \int_0^t \cos(\omega(t-2s)) - \cos(\omega t)\ ds \\ &= \left[\frac{-1}{4\omega} \cdot \sin(\omega(t-2s))\right]_0^t - \left[\frac{s}{2} \cdot \cos(\omega t)\right]_0^t \\ &= \frac{-1}{4\omega} \cdot \sin(-\omega t) - \frac{-1}{4\omega} \cdot \sin(\omega t) - \frac{t}{2} \cdot \cos(\omega t) \\ &= \frac{1}{2\omega} \cdot \sin(\omega t) - \frac{t}{2} \cdot \cos(\omega t). \end{aligned}$$

Substituting this into equation (6.8), yields that for $t \geq 0$,

$$x(t) = \left(\frac{v_0}{\omega} + \frac{a}{2\omega^2}\right) \cdot \sin(\omega t) + \left(x_0 - \frac{at}{2\omega}\right) \cdot \cos(\omega t).$$

Note that by this result, regardless of the amplitude of the resonating oscillation, the last term is unbounded as $t \to \infty$. Therefore, in this case, the particle will spiral out of control whenever the external oscillating force has the same period as that of the particle M without an external force. All of this is exactly in line with the physical expectation of what such situation would result in.

6.4. The Heat Equation

Now it has been made perfectly clear how convolution can be used to solve advanced differential equations, as in the previous two applications. Therefore, let us consider two applications in which this is combined with the Fourier transform, as in section 6.1. The first of these is a fairly easy to solve, but can be used as reference when solving the more complex three-dimensional wave equation in section 6.5. Moreover, this application has a very clear physical interpretation, which will immediately be addressed.

Suppose the x-axis is a heat conductor, and let the temperature at x at time t be given by $u(x,t)$. This function is assumed to be in $\mathcal{E}^2(\mathbb{R})$ with respect to x, and in $\mathcal{E}^1(\mathbb{R})$ with respect to t. Let $c \in \mathbb{R}$ be the heat capacity, for which the amount of heat in 1 unit of the bar is given by $c \cdot u(x,t)$. Furthermore, let γ be the heat conducting coefficient, which can be interpreted as that the amount of heat in the positive-x direction in 1 second is given by $-\gamma \cdot \frac{\partial u}{\partial x}$. Consider some heat sources with heat density $\rho(x,t)$, and assume integrability with respect to both variables. This means that in the interval $\langle x, x+dx\rangle$ and the time interval $\langle t, t+dt\rangle$ the bar receives $\rho(x,t)$ amount of heat. On the other hand, the amount of heat between $\langle x, x+dx\rangle$ during the time interval $\langle t, t+dt\rangle$ is also increased by

$$\left(\gamma \cdot \frac{\partial u}{\partial x}(x+dx,t) - \gamma \cdot \frac{\partial U}{\partial x}(x,t)\right) dt,$$

which is proportional to $\gamma \cdot \frac{\partial^2 u}{\partial x^2}(x,t)\ dx\ dt$. So in total,

$$\left(\gamma \cdot \frac{\partial^2 u}{\partial x^2}(x,t) + \rho(x,t)\right) dx\ dt. \tag{6.9}$$

However, as introduced, the total increase of heat in those intervals should be equal to $c \cdot \frac{\partial u}{\partial t}(x,t)\ dx\ dt$. Combining this with the observation leading to the expression in equation (6.9), gives the so-called heat equation for one dimension:

$$c \cdot \frac{\partial u}{\partial t} - \gamma \cdot \frac{\partial^2 u}{\partial x^2} = \rho. \tag{6.10}$$

Now, as usual let the initial condition be given by $u(x,0) \equiv u_0(x)$, let $\boldsymbol{u}(x,t) \equiv H(t) \cdot u(x,t) \in \mathcal{D}'_+$ and $\boldsymbol{\rho}(x,t) \equiv H(t) \cdot \rho(x,t) \in \mathcal{D}'_+$. Note that since only the time component is to be taken positive,

$$\begin{aligned}\frac{\partial^2 \boldsymbol{u}}{\partial x^2} &= H \cdot \frac{\partial^2 u}{\partial x^2}\\ \frac{\partial \boldsymbol{u}}{\partial t} &= H \cdot \frac{\partial u}{\partial t} + u_0 \cdot \delta.\end{aligned}$$

Therefore, substituting this into the heat equation gives

$$\frac{\partial \boldsymbol{u}}{\partial t} - \frac{\gamma}{c} \cdot \frac{\partial^2 \boldsymbol{u}}{\partial x^2} = \frac{1}{c} \cdot \boldsymbol{\rho} + u_0 \cdot \delta.$$

Taking the Fourier transform with respect to x of both sides, denoted ' $\widehat{\ }$ ' nonetheless, yields

$$\frac{\partial \widehat{\boldsymbol{u}}}{\partial t} + \frac{4\pi^2 \gamma y^2}{c} \cdot \widehat{\boldsymbol{u}} = \frac{1}{c} \cdot \widehat{\boldsymbol{\rho}} + \widehat{u_0} \cdot \delta,$$

recalling that $\mathcal{F}_x(\frac{\partial^2 \boldsymbol{u}}{\partial x^2}(x,t)) = -4\pi^2 y^2 \cdot \mathcal{F}_x \boldsymbol{u}(y,t)$. Then consider the differential operator $D = \frac{\partial}{\partial t} + \frac{4\pi^2\gamma y^2}{c}$, and note that as in Example 4.19, $E \equiv (D\delta)^{-1} = \frac{H}{c} \cdot e^{-\frac{4\pi^2 \gamma y^2}{c} t}$. Fixing y in order to get a convolution equation with respect to t, then gives

$$\widehat{\boldsymbol{u}}(y,t) = \widehat{\boldsymbol{p}} * \left(\frac{H}{c} \cdot e^{-\frac{4\pi^2\gamma y^2}{c} t}\right) + u_0 \cdot H \cdot e^{-\frac{4\pi^2\gamma y^2}{c} t}. \tag{6.11}$$

Taking the inverse Fourier transform would then yield the solution, namely that for $t \geq 0$, $u(x,t) = \bar{\mathcal{F}}_x \widehat{\boldsymbol{u}}(x,t)$. However, this is difficult to determine explicitly in general, but doable for the two special cases presented below. As a final note, it is important to realize how taking the Fourier transform simplified the equations, and allowed us to even compute such general solution.

Special Case: Initial Single Heat Source

Consider the following special case, namely that only initially a single heat source provides a unit of heat to the bar. Thus, the heat density of this source is given by $\rho(x,t) = \delta(x,t)$ Moreover, assume that at the start, no heat is present in the entire bar, i.e. $u_0(x) = 0$. Then, note that $\widehat{\boldsymbol{\rho}}(y,t) = \delta(t)$, and thus by equation (6.11),

$$\widehat{\boldsymbol{u}}(y,t) = \frac{H}{c} \cdot e^{-\frac{4\pi^2\gamma y^2}{c}t}.$$

Hence, by the second case mentioned in Example 5.2,

$$\begin{aligned}
\boldsymbol{u}(x,t) &= \frac{H}{c} \cdot \frac{1}{2\sqrt{\frac{\gamma}{c} \cdot \pi t}} \cdot e^{-\frac{cx^2}{4\gamma t}} \\
\Rightarrow u(x,t) &= \frac{1}{2\sqrt{c\gamma\pi t}} \cdot e^{-\frac{cx^2}{4\gamma t}} \text{ for } t \geq 0. \qquad (6.12)
\end{aligned}$$

As predicted, as time becomes large, the amount of heat in the bar exponentially decreases, and quadratically faster as $|x| \to \infty$, i.e. when moving away from the origin.

Special Case: No Heat Source

A second special case is that when there are no heat sources at all. However, let us not restrict the initial heat distribution of the bar, and thus let us only set $\rho(x,t) = 0$. Then the differential equation (6.10) becomes the so-called diffusion equation, and substituting this into equation (6.11), gives

$$\widehat{\boldsymbol{u}}(y,t) = u_0 \cdot H \cdot e^{-\frac{4\pi^2\gamma y^2}{c}t}.$$

Then, as in the previous special case,

$$\begin{aligned}
\boldsymbol{u}(x,t) &= u_0 * \left(\frac{1}{2\sqrt{\frac{\gamma}{c} \cdot \pi t}} \cdot e^{-\frac{cx^2}{4\gamma t}} \right) \\
u(x,t) &= \left(\frac{u_0}{c}\right) * \left(\frac{1}{2\sqrt{c\gamma \cdot \pi t}} \cdot e^{-\frac{cx^2}{4\gamma t}} \right) \text{ for } t \geq 0.
\end{aligned}$$

Note that this result is very similar to equation (6.12), which was to be expected since both scenarios are very similar. However, it is important here to realize that even though in both examples different parts of equation (6.11) vanished, a very similar result has been derived.

6.5. The Wave Equation

As a final example, let us consider the wave equation in three dimensions. There are numerous ways to set this up with a physical interpretation, but since no physically concrete result or special case will be derived, let us focus on the mathematics behind this. Let us only note that the equation, as suggested in the name, describes the propagation of some wave through some medium, the exact interpretation of which comes down to defining the meaning of the functions and variables involved. As a final note of this, the constant $k > 0$ in this equation can be seen as the maximum propagation speed of such wave, and this will be emphasized in the mathematical result as well.

For now, let $u(x,t)$ be a function that is assumed to be in $\mathcal{E}^2(\mathbb{R}^3)$ with respect to x, and in $\mathcal{E}^2(\mathbb{R})$ with respect to t. Furthermore, since the propagation of the wave may never exceed the speed of light, the solution will always be in the positive light cone. Therefore $T_u \in \mathcal{D}(\mathbb{E}^4)$, as defined in Example 4.15. Then, recall that the Laplace operator 'Δ' is given by

$$\Delta = \sum_{i=1}^{3} \frac{\partial^2}{\partial x_i^2}.$$

Moreover, let $u_0(x) \equiv u(x,0)$ and $v_0(x) \equiv \frac{\partial u}{\partial t}|_{t=0}$. Then the wave equation is given by

$$\frac{\partial^2 u}{\partial t^2} = k^2 \cdot \Delta u. \tag{6.13}$$

Now, note that since differentiation and integration of different variables are interchangeable, $\frac{\partial}{\partial t}\mathcal{F}_x = \mathcal{F}_x \frac{\partial}{\partial t}$. A direct result of this is that $\frac{\partial}{\partial t}\widehat{u}_0(y) = \widehat{v}_0(y)$. Moreover, applying this to equation (6.13), gives

$$\frac{\partial^2 \widehat{u}}{\partial t^2} = -k^2 \cdot |y|^2 \cdot \widehat{u},$$

recalling equation (5.3) once again. This Fourier transformed differential equation is trivial to solve, namely

$$\widehat{u}(y,t) = \widehat{u}_0 \cdot \cos(k|y|\,t) + \widehat{v}_0 \cdot \frac{\sin(k|y|\,t)}{k|y|}. \tag{6.14}$$

Directly computing its inverse Fourier transform is very difficult. However, let us try to work back from a hunch. Namely, consider the so-called Radon distribution $\sigma \in \mathcal{D}'(\mathbb{R}^3)$, given by

$$\langle \sigma, \varphi \rangle \equiv \int_{|x|=1} \varphi(x)\ d\sigma(x) \text{ for } \varphi \in \mathcal{D}(\mathbb{R}^3).$$

Recall that in three dimensions, spherical coordinates for the unit circle are given by

$$(x_1, x_2, x_3) = (\cos\theta_1, \cos\theta_1 \cdot \cos\theta_2, \cos\theta_1 \cdot \sin\theta_2) \text{ for } 0 \le \theta_1 \le \pi, 0 \le \theta_2 \le 2\pi$$

Now, note that for $\sigma(\theta) = (\cos\theta_1, \cos\theta_1 \cdot \cos\theta_2, \cos\theta_1 \cdot \sin\theta_2) = x$, it follows that $\varphi(x) = \varphi(\sigma(\theta)) = \varphi(\cos\theta_1, \cos\theta_1 \cdot \cos\theta_2, \cos\theta_1 \cdot \sin\theta_2)$. Moreover, in this case $d\sigma = \sin\theta_1\ d\theta_1\ d\theta_2$, and thus

$$\langle \sigma, \varphi \rangle \equiv \int_0^{2\pi} \int_0^{\pi} \varphi(\cos\theta_1, \cos\theta_1 \cdot \cos\theta_2, \cos\theta_1 \cdot \sin\theta_2)\ \sin\theta_1\ d\theta_1\ d\theta_2 \text{ for } \varphi \in \mathcal{D}(\mathbb{R}^3).$$

This is the point where it can be used that by Definition 6.1, $\sigma(\theta)$ is radial, and thus

$$\begin{aligned}\widehat{\sigma}(y) = \langle \sigma, e^{2\pi\imath x_1 |y|}\rangle &= \int_0^{2\pi}\int_0^{\pi} e^{2\pi\imath\cos\theta_1|y|}\ \sin\theta_1\ d\theta_1\ d\theta_2 \\ &= 2\pi\cdot\int_0^{\pi} e^{2\pi\imath\cos\theta_1|y|}\ \sin\theta_1\ d\theta_1 \\ &= 2\pi\cdot\left[\frac{-e^{2\pi\imath\cos\theta_1|y|}}{2\pi\imath|y|}\right]_0^{\pi} = \frac{2\pi\cdot\sin|y|}{|y|}.\end{aligned}$$

The result of this computation can be generalized, namely fix $r \geq 0$, and let $\sigma_r \in \mathcal{D}'(\mathbb{R}^3)$ be given by

$$\langle \sigma_r, \varphi\rangle \equiv \int_{|x|=r} \varphi(x)\ d\sigma_r(x) \text{ for } \varphi \in \mathcal{D}(\mathbb{R}^3).$$

Then it follows by a similar computation that

$$\widehat{\sigma}_r(y) = \frac{2\pi r\cdot\sin(r|y|)}{|y|}.$$

Substituting $r = kt$ gives

$$\begin{aligned}\widehat{\sigma}_{kt}(y) &= \frac{2\pi kt\cdot\sin(kt|y|)}{|y|} \\ \Rightarrow \frac{\widehat{\sigma}_{kt}(y)}{2\pi k^2 t} &= \frac{\sin(kt|y|)}{k|y|}.\end{aligned}$$

And thus,

$$\bar{\mathcal{F}}_x\left(\widehat{v}_0(y)\cdot\frac{\sin(kt|y|)}{k|y|}\right) = \frac{\sigma_{kt}(x)}{2\pi k^2 t} * v_0(x). \tag{6.15}$$

Differentiation with respect to t then gives, keeping in mind that differentiation and integration of different variables is interchangeable, and recalling that $v_0(x) = \frac{\partial u}{\partial t}|_{t=0}$,

$$\bar{\mathcal{F}}_x\left(\widehat{w}_0(y)\cdot\frac{\sin(kt|y|)}{k|y|}\right) = \frac{\partial}{\partial t}\left(\frac{\sigma_{kt}(x)}{2\pi k^2 t} * w_0(x)\right),$$

where $w_0(x)$ is unsurprisingly defined as $w_0(x) = \frac{\partial^2 u}{\partial t^2}|_{t=0}$. At this point, there is one thing needed to assume about this application, namely that the initial condition is periodic. Usually, this is the case, as the initial situation can be interpreted as some wave, that will thereafter be propagating through some medium. Specifically, this implies that $w_0(x) = u_0(x)$, and thus

$$\bar{\mathcal{F}}_x\left(\widehat{u}_0(y)\cdot\frac{\sin(kt|y|)}{k|y|}\right) = \frac{\partial}{\partial t}\left(\frac{\sigma_{kt}(x)}{2\pi k^2 t} * u_0(x)\right). \tag{6.16}$$

Taking the inverse Fourier transform of equation (6.14), and then substituting equations (6.15) and (6.16), yields

$$u(x,t) = \frac{\partial}{\partial t}\left(\frac{\sigma_{kt}(x)}{2\pi k^2 t} * u_0(x)\right) + \frac{\sigma_{kt}(x)}{2\pi k^2 t} * v_0(x). \tag{6.17}$$

Computing this convolution product gives

$$\begin{aligned}\sigma_{kt} * u_0 &= \int_{|y|=kt} u_0(x+y)\ d\sigma_{kt}(y) = k^2t^2\cdot\int_{|y'|=1} u_0(x+kty')\ d\sigma(y') \\ &= k^2t^2\cdot\int_0^{2\pi}\int_0^{\pi} u_0(x_1 + kt\cos\theta_1, x_2 + kt\cos\theta_1\cdot\cos\theta_2, x_3 + kt\cos\theta_1\cdot\sin\theta_2)\ \sin\theta_1\ d\theta_1\ d\theta_2 \\ &= k^2t^2\cdot\int_0^{2\pi}\int_0^{\pi} u_0(\psi(\theta))\ \sin\theta_1\ d\theta_1\ d\theta_2, \end{aligned}\tag{6.18}$$

where $(\psi_1, \psi_2, \psi_3) \equiv (x_1 + kt\cos\theta_1, x_2 + kt\cos\theta_1 \cdot \cos\theta_2, x_3 + kt\cos\theta_1 \cdot \sin\theta_2)$, for the sake of brevity. From an almost identical computation, it follows that

$$\sigma_{kt} * v_0 = k^2t^2 \cdot \int_0^{2\pi} \int_0^{\pi} v_0(\psi(\theta)) \ \sin\theta_1 \ d\theta_1 \ d\theta_2. \tag{6.19}$$

Hence, substituting equations (6.18) and (6.19) into (6.17), gives

$$\begin{aligned} u(x,t) &= \frac{\partial}{\partial t}\left(\frac{t}{2\pi} \cdot \int_0^{2\pi} \int_0^{\pi} u_0(\psi(\theta)) \ \sin\theta_1 \ d\theta_1 \ d\theta_2\right) + \frac{t}{2\pi} \cdot \int_0^{2\pi} \int_0^{\pi} v_0(\psi(\theta)) \ \sin\theta_1 \ d\theta_1 \ d\theta_2 \\ &= \frac{1}{2\pi} \cdot \int_0^{2\pi} \int_0^{\pi} u_0(\psi(\theta)) \ \sin\theta_1 \ d\theta_1 \ d\theta_2 + \frac{t}{\pi} \cdot \int_0^{2\pi} \int_0^{\pi} v_0(\psi(\theta)) \ \sin\theta_1 \ d\theta_1 \ d\theta_2 \end{aligned}$$

As a final substitution of variables in this equation, let $(z_1, z_2) \equiv (\cos\theta_1, \sin\theta_1 \cdot \cos\theta_2)$, then

$$dz_1 \ dz_2 = \sin^2\theta_1 \cdot |\sin\theta_2| \ d\theta_1 \ d\theta_2 = \sqrt{1 - |z|^2} \cdot \sin\theta_1 \ d\theta_1 \ d\theta_2.$$

Therefore, the general solution is given by

$$u(x,t) = \frac{1}{2\pi} \cdot \int_{|z|<1} \frac{u_0(x + ktz)}{\sqrt{1 - |z|^2}} \ dz + \frac{t}{\pi} \cdot \int_{|z|<1} \frac{v_0(x + ktz)}{\sqrt{1 - |z|^2}} \ dz. \tag{6.20}$$

As promised, it follows from this solution that $k > 0$ can be seen as the maximum speed of the supposed wave propagation. Namely, consider the case that $u_0(x) = \delta(x)$, $v_0(x) = 0$, then it follows from equation (6.20), that

$$u(x,t) = \left.\frac{1}{2\pi\sqrt{1 - |z|^2}}\right|_{x+ktz=0} = \left.\frac{1}{2\pi\sqrt{1 - |z|^2}}\right|_{z=\frac{-x}{kt}} = \frac{1}{2\pi\sqrt{1 - \frac{x^2}{k^2t^2}}},$$

as long as $\frac{x}{kt} \leq 1$. But that simply implies that k is the maximum speed of such wave propagation. For more concrete special cases, taking a lot of computation, see [9] (Paragraph 5.3).

Finally, note that the involvement of distributions was crucial here, but in a different role than in previous examples. Although the solution defines a distribution in $\mathcal{D}'(\mathbb{E}^4)$, its property of having a unique solution in convolution equations, was not used here (only with respect to the initial conditions). Instead, in order to compute the inverse Fourier transform, a totally different distribution was considered, without even appearing in the final solution. Therefore, also in this last application, it was shown how distributions often play a key role behind the scenes, in numerous ways. This especially is the core observation that one can take away from all these applications, and proves the use of distribution theory.

7. Conclusion and Future Research

As was stated at the very start of this report, the applications in which distribution theory can be used, need not be very difficult. In fact, in the last chapter, a few easy cases have been treated, in which simply stating a differential equation may lead to some very elaborate computations, using some of the more advanced notions of distribution theory. For instance, in order to solve a convolution equation, there is so much foundation of distribution theory needed, to even grasp its meaning. How to solve these equations, how are convolutions even defined, what does it mean for a differential operator to act on a distribution, what space does this solution even live in?

This is one of the main observations to take away from distribution theory, that although an application may look simple, solving it using distributions may involve a lot of advanced computations and theoretical background. And, as had been a recurring theme throughout this report, at the end of the day distributions may not even appear in the derived solution. This is a notion that was first encountered theoretically, when treating test functions and how distributions can be defined using test functions, without any test function occurring in its explicit expression. (As an example of this, recall the observation that $H' = \delta$.)

To this end, the first half of this report lays this foundation of distribution theory, in a very theoretical manner. And as the chapters progress, the examples of this theory become more and more applied. To the end where the chapter covering Fourier transforms, can itself be seen as an application of distribution theory. This also expresses clearly how multi-facaded distributions actually are. First, it was treated how distributions can appear explicitly, by for instance considering the Dirac delta. To reinforce this even further, it has been shown how to differentiate distributions, compute its Taylor extension and extend it to linear forms acting on continuous functions. Then, the notion of distributions became more abstract, even though its applications became more concrete, for instance looking at convolution equations. As was also treated in the last chapter, differential equations in which no distributions initially appear, can be converted to convolution equations in order to derive a solution in which distributions merely play a behind-the-scenes role.

Both of these worlds, the different ways in which distributions can be used, came together when exploring the Fourier transform. Here, it was discovered that distribution theory plays a behind-the-scenes role in solving differential equations by applying the Fourier transform, but distributions were also used explicitly, for instance by computing the Fourier transform of a distribution itself. This duality of distribution theory is an underlying property of distributions that one can take away from this report. Especially in the final chapter, both sides of this coin have been emphasized and worked with.

Of course, there are many more settings in which distribution theory may occur, and also a great number of more advanced ones. For personal research, it may be interesting to explore more of this, and especially look at how distributions are used in Laplace transforms, as distribution theory is even more involved in this than Fourier transforms. In general, it may be a useful suggestion for future research to focus on how distribution theory can be involved more heavily in other mathematical fields. Clearly, there is a very wide range of applications for distributions, and using this in order to clarify patterns or simplify derivations in other fields can be very useful.

Therefore, as a final note, the reader is referred to [1] once more, in order to place their current understanding of distribution theory in perspective with related notions from other fields. This, in turn, may help to see how distributions can be used in those various mathematical fields.

Bibliography

[1] Terrence Tao. *Article on Distributions.* Department of Mathematics, University of California - Los Angeles.

[2] Abdellah El Kinani and Mohamed Oudadess. *Distribution theory and applications*, volume 9. World Scientific Publishing Company, 2010.

[3] Günther Hörmann and Ronald Steinbauer. *Lecture Notes on the Theory of Distributions.* Fakultät für Mathematik, Universität Wien, summer 2009.

[4] Gerrit van Dijk. *Distribution Theory.* De Gruyter Graduate Lectures, De Gruyter, Berlin, 2013.

[5] Erik G. F. Thomas. *Collegedictaat Distributietheorie.* Faculteit der Wiskunde en Natuurwetenschappen, Rijksuniversiteit Groningen, july 1996.

[6] Gustaf Gripenberg. *Fourier Analyysi (Lecture Notes on Fourier Analysis).* Department of Mathematics and Systems Analysis, Aalto University, januari 2009. Translation by Olof Staffans.

[7] Elias M. Stein and Guido Weiss. *Introduction to Fourier analysis on Euclidean spaces*, volume 32. Princeton university press, 2016.

[8] Petre Teodorescu, Wilhelm W Kecs, and Antonela Toma. *Distribution theory: with applications in engineering and physics.* John Wiley & Sons, 2013.

[9] Robert S. Strichartz. *A guide to distribution theory and Fourier transforms.* World Scientific Publishing Company, 1994.

A. Appendix on Definitions

In this report, the notion of integrability is often used. Therefore, let us quickly recap what this exactly entails.

Definition A.1. A function $f : \mathbb{R}^n \to \mathbb{C}$ is called locally integrable if for all compact $K \subset \mathbb{R}^n$

$$\int_K |f(x)| \ dx < \infty.$$

The set of all locally integrable functions on $\mathbb{R}^n$ is denoted L^1, and the norm on this set is defined by

$$\|f\|_1 \equiv \int_{\mathbb{R}^n} |f(x)| \ dx.$$

This notion can be generalized, so for $1 < p < \infty$, a function $f : \mathbb{R}^n \to \mathbb{C}$ is in the set L^p if for all compact sets $K \subset \mathbb{R}^n$

$$\int_K |f(x)|^p \ dx < \infty.$$

Then the so-called L^p-norm on this set is defined accordingly, namely

$$\|f\|_p \equiv \left(\int_{\mathbb{R}^n} |f(x)|^p \ dx \right)^{\frac{1}{p}}.$$

Finally, a norm that is also used regularly is the essential supremum, defined by

$$\|f\|_\infty \equiv \inf \{C > 0 \mid |f(x)| \leq C \ a.e.\},$$

where a.e. stands for 'almost everywhere' and is used in measure theory. The set of functions on which this norm is finite, is denoted L^∞.

B. Appendix on Proofs

In chapter 5, several theorems were encountered for which the proof was left out. Often, this was done because the proof did not add anything to the understanding of the reader. However, a few proofs were simply too long, or would have drawn attention from the main point. Those are the theorems which will be proved below.

The first of which is Theorem 5.8, regarding the derivative of a Fourier transform. Let us repeat the statement of the theorem below.

Theorem B.1. Let $f \in L^1(\mathbb{R})$ and suppose $g(x) \equiv -2\pi\imath x \cdot f(x) \in L^1(\mathbb{R})$. Then $\hat{f} \in \mathcal{E}^1(\mathbb{R})$ and $(\hat{f})' = \hat{g}$.

Proof. By definition of differentiability, at any point $y_0 \in \mathbb{R}$

$$(\hat{f})'(y_0) = \lim_{h\to 0} \frac{\hat{f}(y_0 + h) - \hat{f}(y_0)}{h} = \lim_{h\to 0} \int_{-\infty}^{\infty} f(x)\; e^{-2\pi\imath x y_0} \cdot \left(\frac{e^{-2\pi\imath x h} - 1}{h}\right)\, dx.$$

Note that the absolute value of the integrand can be bounded, namely

$$\left| f(x) \cdot e^{-2\pi\imath x y_0} \cdot \left(\frac{e^{-2\pi\imath x h} - 1}{h}\right)\right| \leq |f(x)| \cdot \left|\frac{e^{-2\pi\imath x h} - 1}{h}\right|. \tag{B.1}$$

Let us bound the second term separately, and to this end note that

$$\left|\frac{e^{-2\pi\imath x h} - 1}{h}\right| = \frac{2 - 2\cos(2\pi x h)}{h} \geq 0,$$

which has a maximum at

$$\begin{aligned}\frac{d}{dh}\left(\frac{2 - 2\cos(2\pi x h)}{h}\right) = \frac{2\pi x \sin(s\pi x h) \cdot 2h - 2 + 2\cos(2\pi x h)}{h^2} &\overset{\text{set}}{=} 0 \\ \Rightarrow 2\pi x h \cdot 2\sin(2\pi x h) + 2\cos(2\pi x h) &= 2 \\ \Rightarrow 2\pi x h \cdot 2\sin(2\pi x h) &= 2 - 2\cos(2\pi x h).\end{aligned}$$

Substituting this last expression into equation B.1 gives a further bound, namely

$$|f(x)| \cdot \left|\frac{e^{-2\pi\imath x h} - 1}{h}\right| \leq |f(x)| \cdot \left|\frac{2\pi x h \cdot 2\sin(2\pi x h)}{h}\right| \leq |f(x)| \cdot |2\pi x| \cdot 2 = 2 \cdot |g(x)|.$$

Therefore, since $g \in L^1(\mathbb{R})$, $|(\hat{f})'(y_0)| \leq 2 \cdot \|g\|_1 < \infty$. Hence, by the dominated convergence theorem, $\hat{f}$ is continuously differentiable, and clearly $(\hat{f})' = \hat{g}$. □

The second theorem that has been left out due to its length, is the important result that the Fourier transform of any distribution with compact support, can be written as the regular distribution of a slowly increasing function. Let us restate Theorem 5.20 below.

Theorem B.2. For any $T \in \mathcal{E}'(\mathbb{R})$, $f(x) \equiv \langle T, e^{-2\pi\imath x y}\rangle$ defines the Fourier transform of T, i.e. $\hat{T} = T_f$.

Proof. By Theorem 3.8, there exists some $\alpha \in \mathcal{D}(\mathbb{R})$ such that $\alpha(x) = 1 \ \forall x \in \operatorname{Supp} T$ and $0 \leq \alpha(x) \leq 1 \ \forall x \in \mathbb{R}$. Then by Proposition 3.21, there exist some $m \in \mathbb{N}$ and $C' > 0$ such that

$$|\langle T, \varphi \rangle| \leq C' \cdot \sum_{k=0}^{m} \sup |\varphi^{(k)}(x)|$$

for all $\varphi \in \mathcal{D}(\mathbb{R})$ satisfying $\operatorname{supp} \varphi \subset \operatorname{supp} \alpha$. Note that by the choice of $\alpha \in \mathcal{D}(\mathbb{R})$, for any $\psi \in \mathcal{E}(\mathbb{R})$ there exists some $\varphi \in \mathcal{D}(\mathbb{R})$ such that $\varphi = \alpha \cdot \psi$, and thus $\varphi(x) = \psi(x) \ \forall x \in \operatorname{Supp} T$. Therefore, for some $C > 0$,

$$|\langle T, \varphi \rangle| = |\langle T, \psi \rangle| \leq C \cdot \sum_{k=0}^{m} \sup |\psi^{(k)}(x)| \text{ for } \psi \in \mathcal{E}.$$

Specifically, it follows that for any fixed $y \in \mathbb{R}$,

$$|\langle T, e^{-2\pi\imath xy} \rangle| \leq C \cdot \sum_{k=0}^{m} \sup_{x \in \mathbb{R}} |(-2\pi\imath xy)^k \cdot e^{-2\pi\imath xy}| = C \cdot \sum_{k=0}^{m} (2\pi)^k \cdot \sup_{x \in \mathbb{R}} |x \cdot y|^{k}.$$

Clearly, for $|x|$ sufficiently large, this expression is bounded by $D \cdot |x|^{m}$ for some $D > 0$. Hence $f(x) = \langle T, e^{-2\pi\imath xy} \rangle$ is slowly increasing and thus, $T_f \in \mathcal{S}'(\mathbb{R})$. This allows us to compute

$$\begin{aligned} \langle \hat{T}, \varphi \rangle = \langle T, \hat{\varphi} \rangle &= \langle T, \int_{-\infty}^{\infty} \varphi(x) \ e^{-2\pi\imath xy} \ dx \rangle \\ &= \langle T, \int_{-\infty}^{\infty} \alpha(y) \ \varphi(x) \ e^{-2\pi\imath xy} \ dx \rangle \\ &= \langle T \otimes T_1, \alpha(y) \cdot \varphi(x) \cdot e^{-2\pi\imath xy} \rangle \\ &= \int_{-\infty}^{\infty} \varphi(x) \cdot \langle T, e^{-2\pi\imath xy} \rangle \ dx \\ &= \langle f, \varphi \rangle \text{ for } \varphi \in \mathcal{D}(\mathbb{R}), \end{aligned}$$

where the fourth line follows from Theorem 4.10. Hence $\hat{T} = T_f$. □