Some Methods In Functional And Harmonic Analysis For Frequency Domain Study Via Fourier Transforms

a dissertation presented
by
Frederico A. C. L. Marinho
to
The Department of Electrical and Electronic Engineering

in partial fulfillment of the requirements
for the degree of
Bachelors
in the subject of
Electrical Engineering

Universidade Federal de Minas Gerais
Belo Horizonte, Minas Gerais
December 2022

Thesis advisor: Professor Pablo D. C. Correa
Thesis co-advisor: Professor Luciano A. F. Santos Frederico A. C. L. Marinho

Some Methods In Functional And Harmonic Analysis For Frequency Domain Study Via Fourier Transforms

Abstract

In this work, the objective is to use functional analysis and harmonic analysis techniques to generalize the Fourier Transform and similar tools in different contexts and levels of abstraction, thus obtaining a better understanding of the frequency domain worked, whatever it may be. The first chapter deals with Fourier Series in Hilbert Spaces and Fourier Transforms in Abelian Locally Compact Topological Groups. Fourier series in Hilbert Spaces converge to an element of the space, which is usually interpreted as a L^2 space, and, therefore, its elements are equivalence classes of functions, therefore it makes no sense to take values in the set of measure space in which it is defined. The Fourier Transforms in Groups seek precisely to solve this part of the problem, making it possible to evaluate an element of space. Then the Theory of Distributions is developed, which allows constructions such as the weak derivative of functions and a generalization of the Fourier transform for mathematical objects that generalize usual functions and encompasses several useful tools such as the Dirac Delta. Then, in the next chapter, Fourier Transforms for Temperate distributions are developed. Finally, the Wavelet Transform Theory is developed, whose main objective is to try to circumvent the Heisenberg uncertainty principle, obtaining a good frequency resolution for low frequencies and a good temporal resolution for high frequencies.

Contents

Listing of figures

This is dedicated to my parents, Frederico and Oriana.

Acknowledgments

I would like to thank every one that supported me on my educational journey, but that would be virtually impossible given that so many people I have met in this years taught me something.

Therefore, I first should thank my family for the entire support I have received along all my live, providing not only the best conditions possible for me to study, but contributing to me in every way they could.

In the department of mathematics, I should thank mainly three professors: Sylvie Marie Oliffson Kamphorst Leal da Silva, Sônia Pinto de Carvalho and Pablo Daniel Carrasco Correa. Sylvie was the first one to welcome me to work in the department of mathematics back when I was taking the first Calculus course at the University. From that moment on I have dedicated a good part of my time taking courses at the department of mathematics and meet many professors that helped me and taught me a lot. Everything started there and I will always be grateful for that opportunity. Sônia after that also started giving me support and helping me whenever I needed, and I needed a lot. Three years of my introductions to scientific research at the University were with Sylvie and Sônia, providing me a good introduction to many subjects that I did not even knew the existence of and that ended up being extremely important in my formation. Pablo also gave me a lot of support during these years since I took a firs course in Dynamical Systems at the second semester. Pablo is also the advisor of this thesis and a person whom, without the support, this text would not exist.

In the department of Electrical and Eletronic Engineering, I also would like to acknowledge three professors: Luciano Antonio Frezzatto Santos, Eduardo Mazoni Andrade Marçal Mendes and Luciano Cunha de Araújo Pimenta. Luciano Frezzatto is the co-advisor of this thesis and helped me a lot with useful suggestions and insights. also making a detailed review of this text. Eduardo also helped me with good suggestions and pointing out some mistakes I made along the way. Luciano Pimenta gave me the support at the end of my bachelors course when I decided to take the Robotics's certificate. He never held me back from taking the classes I was most interested of and that made a huge difference for me.

Finally, I would like to thank two of my friends and colleagues, who reviewed some parts of my text and helped me at specific points of the discussion: Renzo Scarpelli Cabral de Bragança and Gabriel Teixeira Lara Chaves.

It is not knowledge, but the act of learning, not possession, but the act of getting there, which grants the greatest enjoyment.

Carl Friedrich Gauss

0 Introduction

In this text, the main focus is to provide an introduction to some methods in Harmonic and Functional Analysis to study Frequency Domains using Fourier Transforms. This will be done in different levels of abstraction and in various distinct contexts.

It is important to formalize the study of these tools and the correlation between them since the Fourier Analysis revolutionized numerous areas in sciences, mathematics and engineering. It would be virtually impossible to condense all the important results developed since the beginning of this study with Jean-Baptiste Joseph Fourier (1768 – 1830). This texts intends, therefore, to present four different views of this theory accordingly to the contemporary mathematics and the more sophisticated analysis that refined this theory. Some examples and context explanations will be done both in subjects more strictly related to pure mathematics and also in subjects concerning electrical engineering.

This text is divided in four chapters, however, there will be presented essentially four different Fourier Analysis, two of them in Chapter 1, whilst Chapters 2 and 3 present another, and, finally, Chapter 4 presents the last one.

Chapter 1 starts presenting a topic in Functional Analysis concerning Fourier Series in Hilbert Spaces. Although Hilbert Spaces are already a restricted class of vector spaces, this is still way more general and abstract then the Fourier Transform. This starts with some simple facts such as proving that (assuming the Axiom of Choice or Zorn's Lemma) that every Hilbert Space has a very special type of basis that is called a Hilbert basis. With these basis, some of the most natural intuitions that holds in finite dimensional linear spaces still remain true in these more general, possibly infinite dimensional, linear spaces. For example, an orthonormal set will be a basis if, and only if, the Perp Set contains only the zero vector.

After the basic definitions, the Fourier Series arises and are defined for every orthonormal subset $A = \{e_\alpha\}_{\alpha \in J}$ in a Hilbert Space $\mathcal{H}$. In particular, if this set is a Hilbert basis, the Fourier Series converges to some element of the Hilbert space. However, a important Theorem, called the Riesz-Fischer Theorem, will show that every Hilbert space is an ℓ^2 space. And indeed, the Hilbert Spaces are interpreted this way, *i.e.*, generally in many contexts of analysis, the working space is an L^2, and thus consisting of equivalence classes of functions. Therefore, it does not make sense to evaluate some element of the space in a class of functions. To remedy a bit of this situation, a theory concerning Fourier Analysis on groups is a abstract harmonic analysis that provides conditions and interesting results for the Fourier Theory work in much more generality and for one to have a continuous representative in which elements of the space can in fact be evaluated.

Chapter 2 will be dedicated to construct a introduction to the theory of Distributions, which are mathematical objects that describe "generalized functions" such as the Dirac's Delta, widely used in mathematics, physics, engineering and co-related areas. This objects will be linear functionals defined on a space $\mathcal{D}$ that forms the class of test function C_c^∞, and will form a space $\mathcal{D}'$. Many examples are given to motivate this theory. After that, many important topics such as the regular distributions, the duality pairing, pullbacks, etc. Density results are proved and discussed such as the density of $\mathcal{D}$ in L^p and the density of the distributions induced by functions in $\mathcal{D}$ in the space $\mathcal{D}'$. At the end of the Chapter, other tools that are developed such as the derivatives of distributions and the construction of mollifiers that will help in the proof of the density of regular distributions.

In Chapter 3 the development of the theory of distributions continues. However, now another type of distributions and test functions are defined: the test functions of slow growth, forming the space $\mathcal{S}$, and the tempered distributions, forming the space §′. This is done because the Fourier Transform can not be defined on all space $\mathcal{D}$. Functions of slow growth generalize the usual test functions, and thus, the continuous linear functionals on § will also be continuous linear functionals in $\mathcal{D}$, *i.e.*, $\mathcal{S}' \subseteq \mathcal{D}'$. Then, the theory of Fourier Transforms is defined in the spaces L^1, L^2 and $\mathcal{S}$. All of them are compared and co-related in the text. At the end, an ingenious Theorem, called the Riesz-Thorin's Interpolation, is proved and used to develop the theory of Fourier Transforms in L^p for all $1 \leq p \leq 2$ and to prove useful inequalities such as the Hausdorff Young Inequality and Young's Inequality.

Chapter 4 starts with a problem intrinsic to the Fourier Transform: the trade-off between time resolution and frequency resolution. To try to solve this problem, the first tool presented was the Short Time Fourier Transforms (STFT), win which one dives the space into small intervals and does the Fourier Transform as each one. However, this does not solve the problem completely, as one still has, it the time-frequency domain, time intervals for frequency intervals. Moreover, it turns out that a function f and its Fourier Transform $\hat{f} = \mathcal{F}[f]$ can not be both compactly supported and, more over, there is a strong result, called the Heisenberg's Incertainty Priciple that says that the product of the variances of the function and its Fourirer Transform is lowerbounded, and, therefore, it is impossible to increase time resolution without decreasing frequency resolution. The strategy then, was to use multiple time-frequencies resolutions to analyse a signal, varying from parts with bad time resolution and good frequency resolution and parts with good time resolution and bad frequency resolution. This will be done instead of using cosines and sine, using so called wavelets, which in essence are "small waves" with finite spectral power. There will be a mother wavelet ψ from which all others

$$\psi_{a,b} = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right).$$

in a family of wavelets will be generating using two parameters, one for scale a other for translation. This is very useful in engineering for example because, in numerous areas as signal processing, telecommunications and control, one may wish to obtain as much information as possible of the analysis of a signal, and then would seek for high frequency burst in a signal which usually occurs at short time intervals, but also wish to obtain good frequency resolution in other time intervals of the signal. Therefore, a Wavelet Transform

$$W_\psi[f](a,b) = \int_{\mathbb{R}} f(t) \cdot \psi^*_{a,b}(t)\, dt,$$

will be defined and formalized in the resemblance of the Fourier Transform. In this chapter, a specific section will be dedicated to apply this theory to analyse a chirp signal with high frequency interjections.

It is also important to say that some parts of the theory developed in this text may seem repetitive. This will sometimes happen because indeed many of the results are the same but in different contexts and concerning distinct mathematical objects even though the idea is the same. Another reason for this to be the case is that, intentionally in order to give more independence between the chapters, some results are reinforced and recapitulated when necessary. This will be the case, for example, at the beginning of the Chapter 4, when some results about Hilbert and Banach Spaces are presented again to support the necessary theory in this chapter.

Nobody ever figures out what life is all about, and it doesn't matter. Explore the world. Nearly everything is really interesting if you go into it deeply enough.

Richard Feynman

1

Fourier Analysis On Hilbert Spaces And Locally Compact Abelian Groups

1.1 Fourier Analysis In General Hilbert Spaces

Many of the results given in this section can be found in [3].

Recall that a Hilbert Space is a inner product vector space that is a complete metric space with respect to the distance defined by the inner product. That being said, one basic construction is to formulate the notion of a basis for these spaces.

Definition 1. *Let $\mathcal{H}$ be a inner product space and $A \subset \mathcal{H}$ such that for every $u, v \in A$ with $u \neq v$, $\langle u, u\rangle = 1$ and $\langle u, v\rangle = 0$. Then A is said to be a **orthonormal** subset of $\mathcal{H}$.*

With the definition above it is possible to make a precise definition of a special set of elements that form a basis for $\mathcal{H}$.

Definition 2. *A **Hilbert Basis** is a orthonormal subset β of $\mathcal{H}$ such that $\overline{span(\beta)} = \mathcal{H}$.*

It is not entirely obvious that every Hilbert space has a Hilbert basis. Nonetheless, that is indeed the case. However, an important fact about Hilbert Spaces will be required to prove this fact: the *existence of the orthogonal complement*. This existence is a direct consequence of the following lemma:

Lemma 1. *Let $C \subset \mathcal{H}$ be a nonempty closed convex subset, i.e., $C = \overline{C}$ and for every pair of elements $c_1, c_2 \in C$, the line segment $\overline{x_1 x_2}$ connecting x_1 to x_2 is in C. Then, given $x \notin C$, there exists a unique element $y \in C$ such that $x - y \perp C$ and $d(x, y) = d(x, C)$.*

Proof. Without loss of generality, it can be assumed that $x = 0$, since translation is an isometry. Hence, $x \notin C \implies \alpha := d(x, C) > 0$. Take a sequence of elements $(y_n)_{n\in\mathbb{Z}_+} \subseteq C$ such that $d(x, y_n) \xrightarrow{n\longrightarrow\infty} \alpha$. The claim is that (y_n) is a Cauchy sequence. In fact, since $\mathcal{H}$ is an inner product space, the **parallelogram law** holds:

$$\|a - b\|^2 + \|a + b\|^2 = 2(\|a\|^2 + \|b\|^2).$$

Given then y_n, y_m define $z = (y_n + y_m)/2$. Since C is convex, $z \in C$. Therefore:

$$\left\|\frac{y_n - y_m}{2}\right\|^2 = \frac{\|y_n\|^2}{2} + \frac{\|y_m\|^2}{2} - \left\|\frac{y_n + y_m}{2}\right\|^2 \xrightarrow{n,m\longrightarrow\infty} 0,$$

because $\dfrac{\|y_n\|^2}{2} \xrightarrow{n\longrightarrow\infty} \dfrac{\alpha^2}{2}$, $\dfrac{\|y_m\|^2}{2} \xrightarrow{m\longrightarrow\infty} \dfrac{\alpha^2}{2}$ and $\left\|\dfrac{y_n + y_m}{2}\right\|^2$ is a number greater or equal to α^2. Hence the existence of $y \in C$ such that $d(x, y) = d(x, C)$ is proved. To show that y is unique, suppose that there exists $y' \neq y$ satisfying the same conditions. Then:

$$\left\|\frac{y - y'}{2}\right\|^2 = \frac{\|y\|^2}{2} + \frac{\|y'\|^2}{2} - \left\|\frac{y + y'}{2}\right\|^2 \leq 0,$$

due to a similar argument as the one in the previous parallelogram inequality. Therefore, $y = y'$, a contradiction. The fact that $y - y \perp C$ is a direct consequence of Pythagoras's Theorem. □

Corollary 1. *Given a closed subspace $C \subseteq \mathcal{H}$, there exists another closed subspace $C^\perp$ such that $C \oplus C^\perp$. The set $C^\perp$ is known as the **Orthogonal Complement** of C.*

Proof. Straightforward from Lemma 1. □

Proposition 1. *Every Hilbert space $\mathcal{H}$ has a Hilbert basis.*

Proof. Let $\mathcal{F}$ be the family of all orthonormal subsets of $\mathcal{H}$ endowed with the partial order of inclusion, considering, for $F_1, F_2 \in \mathcal{F}$:

$$F_1 \leq F_2 \iff F_1 \subseteq F_2.$$

For a given totally ordered chain $\{F_i\}_{i\in I}$ in $\mathcal{F}$, it is clear that $\cup_{i\in I} F_i$ is an upper bound in $\mathcal{F}$ for $\{F_i\}_{i\in I}$. From Zorn's Lemma, there is a maximal element $\mathcal{M} \subseteq \mathcal{H}$ such that $\mathcal{M}$ is orthonormal.

If $\overline{\operatorname{span}(\mathcal{M})} = \mathcal{H}$, the result is proven. Otherwise, suppose $\overline{\operatorname{span}(\mathcal{M})} \neq \mathcal{H}$. Then

$$\mathcal{H} = \overline{\operatorname{span}(\mathcal{M})} \oplus \overline{\operatorname{span}(\mathcal{M})}^\perp.$$

If there exists $u \in \overline{\text{span}(\mathcal{M})}^{\perp}$ with $\|u\| = 1$, then

$$\mathcal{M} \cup \{u\} \supsetneq \mathcal{M} \implies \mathcal{M} \cup \{u\} > \mathcal{M},$$

contradicting the fact that $\mathcal{M}$ is maximal. □

An equivalent verifying that a subset is a Hilbert basis is given by the following result:

Proposition 2. *Let β be an orthonormal subset of $\mathcal{H}$. The following claims are equivalent:*

1. *β is a Hilbert Basis for $\mathcal{H}$.*
2. *$\beta^{\perp} = \{0\}$.*

Proof. (1) $\implies$ (2) : Let $v \in \mathcal{H}$ and $\varepsilon > 0$. By Proposition 1, there is a Hilbert basis $\beta = \{e_\lambda \mid \lambda \in \Lambda\}$ for $\mathcal{H}$. Therefore, for some finite subset $\{e_{\lambda j} \mid \lambda j \in \Lambda, j \in \{1, \ldots, m\}\}$ of span() it is possible to write

$$\left\| x - \sum_{i=1}^{m} a_{\lambda j} e_{\lambda j} \right\| < \varepsilon.$$

If $x \in \beta^{\perp}$, then, without loss of generality supposing $\varepsilon < 1$:

$$\varepsilon > \left\| x - \sum_{j=1}^{m} a_{\alpha j} e_{\alpha j} \right\|^2 \overset{\star}{=} \|x\|^2 + \left\| \sum_{j=1}^{m} a_{\alpha j} e_{\alpha j} \right\|^2 \geq \|x\|^2,$$

where $(\star)$ is valid because $x \in \beta^{\perp}$. Thus, $\|x\| < \varepsilon$ for every $\varepsilon > 0$, implying that $x = 0$.

(2) $\implies$ (1) : Let $M = \overline{\text{span}(\beta)}$ and write $\mathcal{H} = M \oplus M^{\perp}$. Assuming condition (2), $M^{\perp} \subseteq \beta^{\perp} = \{0\}$ and $\{0\} \in M^{\perp}$. Hence, $M^{\perp} = \{0\}$, implying that $\mathcal{H} = M$, *i.e.*, β is a Hilbert basis. □

Recall that for a summation over an uncountable index set, we define the sum as being the suprema as the sum runs over all finite subset of indexes. More precisely, if J is an uncountable index set, then

$$\sum_{\alpha \in J} a_\alpha = \sup \left\{ \sum_{\alpha f \in F \subset J} a_{\alpha f} \ : \ |F| < \aleph_0 \right\}$$

In order to define the notion of Fourier Series with respect to an orthonormal subset of $\mathcal{H}$, it's useful to have the following result:

Proposition 3 (Bessel's Inequality). *Let $A = \{e_\alpha \mid \alpha \in J\}$ an orthonormal subset of $\mathcal{H}$. Then, for every $x \in \mathcal{H}$:*

$$\sum_{\alpha \in J} |\langle x, e_\alpha \rangle|^2 \leq \|x\|^2.$$

In particular, $\langle x, e_\alpha \rangle \neq 0$ for only a countable subset $J_x \subset J$ (for each x).

Proof. Fix a countable subset $F = \{e_{\alpha f}\}_{f \in \mathbb{N}}$ of the index set J. Given any $x \in \mathcal{H}$, then

$$y_m := x - \left(\sum_{f=1}^{m} \langle x, e_{\alpha f} e_{\alpha f} \rangle \right),$$

is such that $y_m \perp e_{\alpha f}$ for all $f \in \{1, \ldots, m\}$. Thus,

$$\|x\|^2 = \|y_m\|^2 + \left\| \sum_{f=1}^{m} \langle x, e_{\alpha f} \rangle e_{\alpha f} \right\|^2 = \|y_m\|^2 + \sum_{f=1}^{m} |\langle x, e_{\alpha f} \rangle|^2 \geq \sum_{f=1}^{m} |\langle x, e_{\alpha f} \rangle|^2.$$

Therefore, for any finite subset of indexes of F,

$$\|x\|^2 \geq \sum_{f=1}^{m} |\langle x, e_{\alpha f} \rangle|^2,$$

Since F is countable, then

$$\sum_{\alpha f \in F} |\langle x, e_{\alpha f} \rangle|^2 \leq \|x\|^2. \qquad (\star)$$

So, for a given $x \in \mathcal{H}$, define

$$F_m = \left\{ \alpha f \in F \, : \, |\langle x, e_{\alpha f} \rangle| \geq \frac{1}{m} \right\}.$$

By the relation in $(\star)$, $|F_m| < \aleph_0$ for all $m \in \mathbb{N}$. Therefore,

$$\bigcup_{m \in \mathbb{N}} F_m = \left\{ \alpha f \in F \, : \, \langle x, e_{\alpha f} \rangle \neq 0 \right\},$$

is countable as countable union of finite sets. □

Now it is possible to define, in a formal way at first, the Fourier Series:

Definition 3. *Let $A = \{e_\alpha\}_{\alpha \in J}$ be a orthonormal subset in a Hilbert space $\mathcal{H}$. The set $\{\langle x, e_\alpha \}_{\alpha \in J}$ is called the set of **Fourier Coefficients** of $x \in \mathcal{H}$. Furthermore, the series*

$$\sum_{\alpha \in J} \langle x, e_\alpha \rangle e_\alpha,$$

*is called the **Fourier Series** of x with respect to A.*

Theorem 1. *Let $\beta = \{e_\alpha\}_{\alpha \in J}$ be a orthonormal subset of $\mathcal{H}$. Then, the following statements are equivalent:*

1. *β is a Hilbert basis for $\mathcal{H}$.*
2. *$\forall x \in \mathcal{H}, \; \sum_{\alpha \in J} \langle x, e_\alpha \rangle e_\alpha$. (Fourier Series converges for all $x \in \mathcal{H}$)*
3. *$\forall x \in \mathcal{H}, \; \|x\|^2 = \sum_{\alpha \in J} |\langle x, e_\alpha \rangle|^2$. (Parseval's Identity)*

Proof. Given any $x \in \mathcal{H}$, by Proposition 3, $\langle x, e_\alpha \rangle \neq 0$ only for a countable subset of the index set $S \subset F$. Then, because $|F| \leq \aleph_0$, either F is finite or $F \cong \mathbb{N}$. Suppose, without loss of generality, $F = \mathbb{N}$.

(1) $\implies$ (2) : By Bessel's inequality, $\sum_{j\in S}|\langle x, e_j\rangle|^2$ converges. The claim here is that $\sum_{j\in S}\langle x, e_j\rangle e_j$ converges in $\mathcal{H}$.

If $|S| < \aleph_0$, this fact is trivial. Suppose then that $|S| = \aleph_0$. Then, for $m \geq n$

$$\left\|\sum_{j=n}^{m}\langle x, e_j\rangle e_j\right\|^2 = \sum_{j=n}^{m}|\langle x, e_j\rangle|^2 \xrightarrow{m,n\longrightarrow\infty} 0.$$

Thus, $S_n = \sum_{j=1}^{n}\langle x, e_j\rangle e_j$ is Cauchy and therefore converges as $\mathcal{H}$ is a Hilbert space and is complete by definition.

Hence,

$$\begin{cases} \left\langle \sum_{j\in S}\langle x, e_j\rangle e_j, e_k \right\rangle = \langle x, e_k\rangle, \quad \forall k \in S \\ \\ \left\langle \sum_{j\in S}\langle x, e_j\rangle e_j, e_\alpha \right\rangle = 0, \quad \forall k \notin S. \end{cases}$$

If $y = x - \sum_{j\in S}\langle x, e_j\rangle e_j$, then

$$\langle y, e_\alpha\rangle = \begin{cases} 0, \quad \text{if } \alpha \in S, \\ \\ \langle x, e_\alpha\rangle, \quad \text{if } \alpha \notin S. \end{cases}$$

However, only the elements of S satisfy $\langle x, e_\alpha\rangle \neq 0$. Thus $\langle y, e_\alpha = 0 \forall \alpha \in J$. Therefore, $y \in \beta^\perp = \{0\}$, because β is a Hilbert basis. Finally, one can conclude that

$$x = \sum_{j\in S}\langle x, e_j\rangle e_j.$$

(2) $\implies$ (3) : If $|S| < \aleph_0$, this fact is straightforward. If $|S| = \aleph_0$, then

$$\underbrace{\left\| x - \sum_{j=1}^{m} \langle x, e_j \rangle e_j \right\|^2}_{\xrightarrow{m \longrightarrow \infty} 0} = \|x\|^2 \underbrace{-2\mathfrak{Re}\left(\left\langle x, \sum_{j=1}^{m} \langle x, e_j \rangle e_j \right\rangle\right)}_{-2\sum_{j=1}^{m} |\langle x, e_j \rangle|^2} + \underbrace{\left\| \sum_{j=1}^{n} \langle x, e_j \rangle e_j \right\|^2}_{\sum_{j=1}^{m} |\langle x, e_j \rangle|^2}.$$

Therefore,

$$\|x\|^2 - \sum_{j=1}^{m} |\langle x, e_j \rangle|^2 \longrightarrow 0,$$

i.e.,

$$\|x\|^2 = \sum_{j \in S} |\langle x, e_j \rangle|^2.$$

(3) $\implies$ (1) : Let $x \in \beta$. Then, $\langle x, e_\alpha \rangle$ for all $\alpha \in J$, and, assuming (3), $\|x\| = 0 \implies x = 0$. Thus, $\beta^\perp = \{0\}$ and hence β is a Hilbert basis. □

Definition 4. *A surjective map $\mathcal{H} \to \mathcal{H}_1 = U(\mathcal{H})$, where $\mathcal{H}$ and $\mathcal{H}_1$ are Hilbert spaces is said to be* ***unitary*** *if it is linear and $\langle U(x), U(y) \rangle_{\mathcal{H}_1} = \langle x, y \rangle_{\mathcal{H}}$ for all $x, y \in \mathcal{H}$. In particular, $\|x\|_{\mathcal{H}} = \|U(x)\|_{\mathcal{H}_1}$.*

Definition 5. *Let J be an index set. Define the set $\ell^2(J)$ as*

$$\ell^2(J) = \left\{ x = (x^\alpha)_{\alpha \in J},\ x^\alpha \in \mathbb{K} \ :\ \sum_{\alpha \in J} |x^\alpha|^2 < \infty \right\},$$

where $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$.

Observation 1. *In the next Theorem the space ℓ is defined just as the space L but assuming the counting measure and the sigma algebra of parts.*

Theorem 2 (Riesz-Fisher)**.** *Let $\mathcal{H}$ be an infinite-dimensional Hilbert space. Then, there exists a set J and a unitary operator $U : \mathcal{H} \to \ell^2(J) = U(\mathcal{H})$.*

In particular, if $|J| > \aleph_0$*, then* $x \in \ell^2(J) \implies x^\alpha \neq 0$ *only for a countable subset of indexes.*

Proof. The idea for proving that is the same as what was done for proving Bessel's Inequality.

Indeed, take a Hilbert basis $\beta = \{e_\alpha\}_{\alpha \in J}$ for $\mathcal{H}$ and define:

$$\begin{aligned} U : &\mathcal{H} \to \ell^2(J) \\ &x \mapsto (\langle x, e_\alpha \rangle)_{\alpha \in J} . \end{aligned}$$

As defined, U is linear and a isometry (Parseval's Identity) and therefore U is unitary due to the Polarization Identity. It suffices to show that $U(\mathcal{H}) = \ell^2(J)$.

Let $y \in \ell^2(J)$. Then $J_r = \{\alpha \in J \,:\, y^\alpha \neq 0\}$ by the what was said previous to the theorem. Define

$$\sum_{j \in J_r} y^j e_j.$$

Therefore, this series converges in $\mathcal{H}$. Indeed, taking the partial sums, that is limited by $\|y\|_2$, and hence the series converges as monotone limited sequence.

Define

$$x = \sum_{j \in J_r} y^j e_j.$$

Thus, for every $w \in \mathcal{H}$,

$$\langle x, w \rangle = \left\langle \sum_{j \in J_r} y^j e_j, w \right\rangle = \sum_{j \in J_r} y^j \langle e_j, w \rangle.$$

Therefore, for $w = e_\alpha$,

$$\langle x, e_\alpha \rangle = y^\alpha \iff U(x) = y.$$

□

Corollary 2. *Let* $\mathcal{H}$ *be a Hilbert space and* $A = \{x_\alpha\}_{\alpha \in J}$ *be a orthonormal subset of* $\mathcal{H}$ *(not necessarily* *a basis). Let* $\mathcal{H}_0 = \overline{\mathrm{span}(\mathrm{A})}$ *(a closed subspace of* $\mathcal{H}$ *by definition) and* $\Pi : \mathcal{H} \to \mathcal{H}_0$ *the orthogonal projection in* $\mathcal{H}_0$*. Then*

$$\Pi(x) = \sum_{\alpha \in J} \langle x, x_\alpha \rangle x_\alpha.$$

Proof. Since $\mathcal{H}_0$ is a Hilbert space, there must exists a map $U : \mathcal{H}_0 \to \ell^2(J)$ that is surjective and unitary. Given any $x \in H$,

$$\Pi(x) \in \mathcal{H}_0 \implies \Pi(x) = U^{-1}(y) = \sum_{\alpha \in J} y^\alpha x_\alpha,$$

where $y = (y^\alpha)_{\alpha \in J} \in \ell^2(J)$. Furthermore, $y^\alpha = \langle \Pi(x), x_\alpha \rangle$. Note that $x = \Pi(x) + \Pi_\perp(x)$ and $\langle \Pi_\perp(x), x_\alpha \rangle = 0$ for all α. Hence, $\langle x, x_\alpha \rangle = \langle \Pi(x), x_\alpha \rangle$ in a way that

$$\Pi(x) = \sum_{\alpha \in J} \langle x, x_\alpha \rangle x_\alpha.$$

$\square$

The following result is extremely useful and characterizes an important class of Hilbert spaces in which many theories takes places.

Theorem 3. *Let* $\mathcal{H}$ *be a Hilbert space. The following statements are equivalent:*

1. $\mathcal{H}$ *is separable.*

2. *Every Hilbert basis for* $\mathcal{H}$ *is countable.*

3. *There exists a surjective unitary operator* $U : \mathcal{H} \to \ell^2(\mathbb{N})$ *if* $\dim(\mathcal{H}) \geq \aleph_0$ *or* $U : \mathcal{H} \to \mathbb{R}^m$ *if* $\dim(\mathcal{H}) < \aleph_0$.

Proof.

$(2) \implies (3)$: It is an immediate consequence of Theorem 2.

(3) $\implies$ (1) : This fact is also straightforward because U is unitary and both $\ell^2(\mathbb{N})$ and $\mathbb{R}^m$ are separable.

(1) $\implies$ (2) : Let $\beta \subseteq \mathcal{H}$ be a Hilbert basis for $\mathcal{H}$. Hence, β is separable, since $\mathcal{H}$ is separable. Let $\alpha \subseteq \beta$ be a dense countable subset with $\beta \subseteq \overline{\alpha}$. The claim is that $\alpha = \beta$. Indeed, suppose that this is not true, *i.e.*, $\alpha \subsetneq \beta$, and let $b \in \beta$ such that $b \notin \alpha$. Therefore, for all $a \in \alpha \subsetneq \beta$:

$$\|a - b\|^2 = \|a\|^2 + \|b\|^2 = 2,$$

since β is orthonormal. Hence, $b \notin \overline{\alpha}$, a contradiction.

Hence, $\alpha = \beta$, and β is a countable dense subset. □

One of the main problems in dealing with Fourier analysis on general Hilbert spaces is the fact that, most of the times, the space that is concerned is an L^2 space, *i.e.*, an space of equivalent classes. It is a standard result in measure theory that every $L^p(X, \mathcal{X}, \mu)$ space for $p \geq 1$ is complete an thus is a Banach space. The proof of this fact can be found in [2]. However, in fact, every Hilbert space is equivalent to an L^2 space:

Corollary 3. *Every Hilbert Space is isomorphic to an* $L^2(X, \mathcal{X}, \mu)$ *space.*

Proof. This is a consequence of the Riesz-Fischer Theorem.

□

To finish this section, it is important to discuss the concern in the discussion in dealing with Hilbert space introduced earlier.

As said before, usually the Hilbert space is actually a quotient set of equivalence classes of the Lebesgue square integrable functions. Therefore, the convergence of the Fourier Series is in L^2, *i.e.*, in the norm L^2. Hence, it does not make any sense to evaluate the element that results of the convergence in the Fourier series of the function in the space $L^2(X, \mathcal{X}, \mu)$ in an element of the space X as its value its not defined in every point and the elements of the equivalence class only being equal $\mu - a.e.$

For one to be able to require a convergence and a possible evaluation at a element of the space, new methods and assumptions are required in a new approach to the problem. This will be discussed in the next section.

1.2 Fourier Analysis with Haar Measure on Topological Groups

Let $L^2(X, \mathcal{X}, \mu) = \mathcal{H}$ be a Hilbert space. Here, the separability of the space is a desirable property and that is not always the case. For that to be assured, some properties must be satisfied such as requiring that the space is regular and the Lebesgue σ-algebra is taken.

Suppose that $\mathcal{H}$ is separable. Then, by Theorem 3 there exists a countable orthonormal basis $\{e_n\}_{n\in\mathbb{N}}$. Thus, if $f \in L^2(X, \mathcal{X}, \mu)$:

$$f = \sum_{n\in\mathbb{N}} \langle f, e_n \rangle .$$

It is important to reinforce that here f is not a function but an equivalent class of functions, the same being valid for each e_n.

Consider the circle $\mathbb{T} = S^1 = \mathbb{R}/\mathbb{Z}$ and the set of functions $\{e_n\}_{n\in\mathbb{Z}}$ such that:

$$\begin{aligned} e_n : [0,1) &\longrightarrow S^1 \\ t &\longmapsto e^{int}. \end{aligned}$$

Then $\{e_n\}$ (here again considering equivalent classes) forms an orthonormal basis for $L^2(\mathbb{T})$. In fact, there are infinitely many basis for $L^2(\mathbb{T})$, but this one is special in many ways that will soon be clear.

Definition 6. *A **Topological Group** is a topological space G that is also a group and such that the group operation and the inversion map (that takes each g to its inverse g^{-1}) are continuous.*

Definition 7. *Let G be a topological group that is abelian, locally compact Hausdorff space. A* ***character*** *is a continuous morphism.*

The idea here is analogous to the idea of a dual space of a vector space $V^* = \{T : V \to \mathbb{K} \mid T \text{linear}\}$, for which, in the context of functional analysis, one usually requires the maps to be continuous.

The choice for the group S^1 is due to the fact that S^1 has one of the most simple group structures appropriated for this context. Indeed, S^1 has dimension 1 and is compact in its natural topology (the quotient topology), while the real line $\mathbb{R}$ is not compact.

Definition 8. *Let G be a topological group that is abelian, locally compact Hausdorff space. Define the* ***dual group*** *with respect to S^1 as being the set $\hat{G}$ such that*

$$\hat{G} = \{\xi : \xi \text{ is a character of } G\}.$$

The reader is encouraged to try to proof the following example

Example 1. $\hat{\mathbb{T}} \cong \mathbb{Z}$. *More precisely, if $\xi : \mathbb{T} \to S^1$ is a character, then $\xi = e_n$ for some $n \in \mathbb{Z}$.*

Indeed, a more general result than the one given in Example 1 above is that if G is a compact Hausdorff group, then the dual group $\hat{G}$ is discrete. Moreover, in this case of G Hausdorff compact, the dual group $\hat{G}$ will be countable if and only if G is second countable (and, since G is Hausdorff, will also be metrizable).

Example 2. *It is interesting to note that not every discrete dual group is countable. In fact, there exists uncountable discrete groups as well as there are non-metrizable compact groups. For example, let Λ be an uncountable set and consider the "Tubby-Torus" group:*

$$G = \mathbb{T}^{\Lambda}$$

Consider the usual topology and the usual group structure on $\mathbb{T}$. Define the topology on G as the product topology and the group structure as the canonical product group. Then the dual group can be written as

$$\hat{G} = \bigoplus_{\lambda \in \Lambda},$$

an uncountable direct sum, as a discrete group. The converse can also be done, i.e., given any uncountable discrete group H, then $\hat{H}$ is a non-metrizable compact group.

In the next part is useful to recall the following definition.

Definition 9. *Let (X, τ) be a topological space and $(X, \mathcal{X}, \mu)$ a measure space. Then $(X, \tau, \mathcal{X}, \mu)$ is said to be a **topological measure space**. Furthermore, a measurable subset A of X is said to be **inner regular** if*

$$\mu(A) = \sup \left\{ \mu(K) \;:\; K \subseteq A \text{ is compact} \right\},$$

and $\mu(K) < \infty$ for all compact sets K.

*In a similar way, A is said to be **outer regular** if*

$$\mu(A) = \inf \left\{ \mu(G) \;:\; G \supseteq A \text{ is open} \right\},$$

and $\mu(K) < \infty$ for all compact sets K.

*A measure is called **inner regular** if every measurable set is inner regular (although in some text, the authors require that only that every open measurable set is inner regular). The measure is called **outer regular** if every measurable set is outer regular. A measure is called **regular** if it is both outer regular and inner regular.*

The following measure construction plays a central role in this approach to a more abstract Fourier Analysis.

Theorem 4. *Let G be a topological group that is abelian, locally compact Hausdorff space. Then there exists a measure μ_G on the Borel subsets of G that is both regular and* ***translational invariant****, i.e., considering $L_g : G \to G$, $x \mapsto xg$, then*

$$L_{g_*}\mu_G = \mu_G,$$

where $*$ denotes the pushforward of the measure.

Furthermore, given another regular measure μ satisfying the above conditions, they are equal up to a constant factor (i.e., $\mu = \lambda\mu_G$ for some constant λ and every μ).

Definition 10 (Haar Measure)**.** *The measure described in Theorem 4 is called a* ***Haar Measure****.*

Example 3. *The Lebesgue measure λ is a Haar measure. The same could be said to 2λ.*

Let G be compact (not only locally) and fix a Haar measure μ_G. Considering $L^2(G)$, then $\{\xi \in \hat{G}\}$ is an orthonormal basis $L^2(G)$ (see [9]). This compact requirement is important otherwise functions like 1_G would not be integrable.

It is often convenient to take the **Haar Probability** when the measure is finite, uniquely determined taking

$$\mathbb{P}_G = \frac{\mu_G}{\mu_G(G)}.$$

For a Fourier Series

$$f = \sum_{n\in\mathbb{Z}} a_n e_n,$$

this equality holds in L^2 spaces, which are Hilbert spaces.

It is easy to prove that two separable Hilbert spaces are isomorphic.

Nonetheless, more properties are require for the theory to be consistent. Remember that it does not make sense to evaluate an element of L^2 as it is equivalent class of functions.

In fact, two facts are usually assumed in this type of analysis to make the evaluation of the function possible in some sense. First, one desirable property of the space would be to have the set of continuous functions dense in the space and such that every element of the L^p space considered has only one continuous representative in each equivalent class. The other property is that the measure of the space is positive in open sets in a sense that will be made precise in the following definition.

Definition 11. *A measure is said to have* ***total support*** *if every open set has positive measure.*

It is important in this context for the measure to be positive in open sets. That because if a function h is continuous and

$$\begin{cases} h \geq 0, \\ \displaystyle\int_X h \, d\mu \end{cases}$$

then h will in fact identically 0 and not just 0 almost everywhere.

In fact, let U be an open set. Then $\mu(U) > 0$, and, considering h a non negative measurable function:

$$\int_U h \, d\mu \leq \int_X h \, d\mu \equiv 0,$$

and thus, since h is continuous, if $h(x) > 0$ for some $x \in X$, then $h > 0$ in $B_r(x)$ for some $r > 0$ and the integral in U would be necessarily positive, an absurd.

Now, more conditions are required to evaluate the function representative of the class and require the convergence everywhere (not only almost everywhere), which is a stronger convergence condition. In fact, if f is differentiable, then the series will in indeed converge.

Definition 12. *Let G be a locally compact abelian group and λ_G a Haar measure in G. Then, the* ***Fourier Transform*** *of a function $F : G \longrightarrow \mathbb{C}$ is defined as being the function $\hat{F} : \hat{G} \longrightarrow \mathbb{C}$ given by*

$$\hat{F}(\xi) = \int_G F\bar{\xi} \, d\lambda_G = \int_{g \in G} F(g)\overline{\xi(g)} \, d\lambda_G(g).$$

Moreover, thinking about the Fourier Series of a periodic functions in terms of its characters, the Fourier Coefficients are precisely the transforms:

$$f \sim \sum_{n\in\mathbb{Z}} a_n e_n \implies a_n = f(\omega) = \int f(t) \cdot \underbrace{e^{-i\omega t}}_{\text{characters}} \, dt, \qquad (\omega \in \mathbb{R}).$$

Mathematics is a game played according to certain simple rules with meaningless marks on paper.

David Hilbert

2

Distributions

2.1 Introduction and Motivation For Distributions

In the general context of real analysis, some concepts arises as founding blocks to the theory, such as the notions of functions (here in the naive set theory approach), limits, derivatives and so forth. However, in some other fields of mathematics, such as the study of differential equations and Fourier Transforms, and also in engineering, in areas such as signal processing, telecommunications and classical theory of control, some interesting solutions to many problems are functions with sharp turns or, in a more critical scenario, not functions in the usual way of the definition.

The idea of the concept of some "generalized functions" has its motivations in trying to define derivatives everywhere for functions that are not differentiable in the usual sense.

Example 4. *Consider the Heaviside Step Function $u : \mathbb{R} \to \mathbb{R}$ defined as*

$$u(x) = \begin{cases} 1, & \text{if } x \geq 0; \\ 0, & \text{if } x < 0. \end{cases}$$

Clearly, observing the graph of u in Figure 2.1, it is obvious that u does not have a derivative at 0.

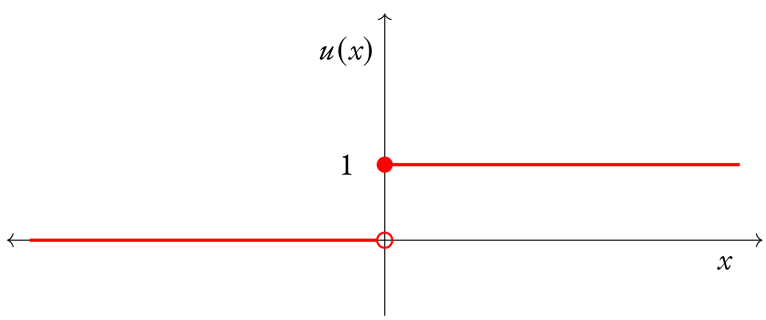

Figure 2.1: Heaviside's Step Function Graphic.

Therefore, in the usual approach, it is not possible to define a derivative u' on all the real line.

The idea of trying to elaborate a "function" that would be the derivative u' lead to the formulation of the Dirac's Delta Function δ (which is not a function).

Since u is constant on both $\mathbb{R}^*_+$ and $\mathbb{R}^*_-$, it is clear that this new "function" must satisfy $\delta(x) = 0$ for every $x \neq 0$. Nonetheless, it is desirable that the F.T.C. still holds in some sense. More precisely, for every $\varepsilon > 0$, it is wanted that

$$\int_{-\varepsilon}^{\varepsilon} \delta(x)\, dx = \int_{-\varepsilon}^{\varepsilon} u'(x)\, dx = u(\varepsilon) - u(-\varepsilon) = 1 \,\forall\, \varepsilon > 0.$$

Hence, the ∂ can be described imposing the following two properties:

$$\begin{cases} \partial(x) = 0 \textit{ for all } x \neq 0; \\ \displaystyle\int_{-\varepsilon}^{\varepsilon} \partial(x)\, dx = 1 \,\forall\, \varepsilon > 0. \end{cases}$$

This "generalized function" would look something like what is shown in Figure 2.2 onward, where the arrow represents the fact that the amplitude of ∂ in 0 should be infinite:

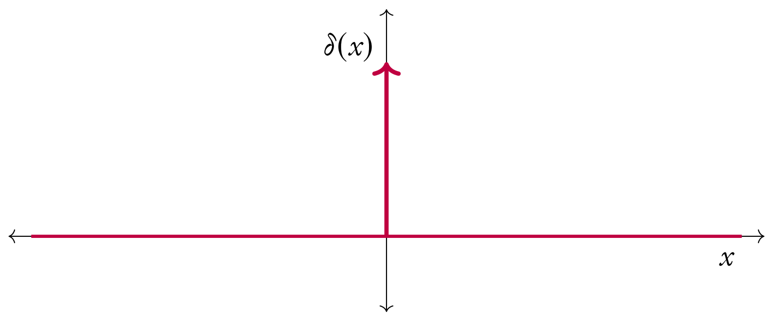

Figure 2.2: Dirac's Delta "generalized function".

The amplitude of the Delta is infinite at 0, however, its weight is said to be equal to 1 as the integral is required to be finite and equal to 1.

Therefore, it is evident that ∂ can not be a function in the usual sense. To be entirely clear, however, the first requirement says that ∂ is 0 almost everywhere. Considering the Lebesgue measure on $\mathbb{R}$, this means that:

$$\int_U \partial \, d\lambda = 0,$$

for every measurable $U \subseteq \mathbb{R}$, which contradicts the second requirement.

The Dirac's Delta can also be defined in $\mathbb{R}^n$ as the product of n Deltas, i.e., if $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, then

$$\partial(x) = \partial(x_1) \cdots \partial(x_n).$$

To finish this example, the case gets even worse when there were attempts to define the higher order derivatives such as δ', δ'', and so forth.

The goal then was to define those "generalized functions" in a more precise way and give a meaning to these derivatives. Motivated by problems as the ones in Example 4, the concept of a new mathematical object was created, the *Distributions*. Latter, Dirac's Delta Function will be defined as a Distribution.

The general idea for the concept of distributions that will be defined later is to think about it as densities. For example, when dealing with theses densities in the context of generalizing real-valued functions of one real variable, these distributions could represent the linear mass density of a one dimensional model of a bar. In the case of the Delta function, the density would be a singular density as all the mass is concentrated in one point. This is illustrated for a function $f : \mathbb{R} \to \mathbb{R}$ in Figure 2.3 ahead.

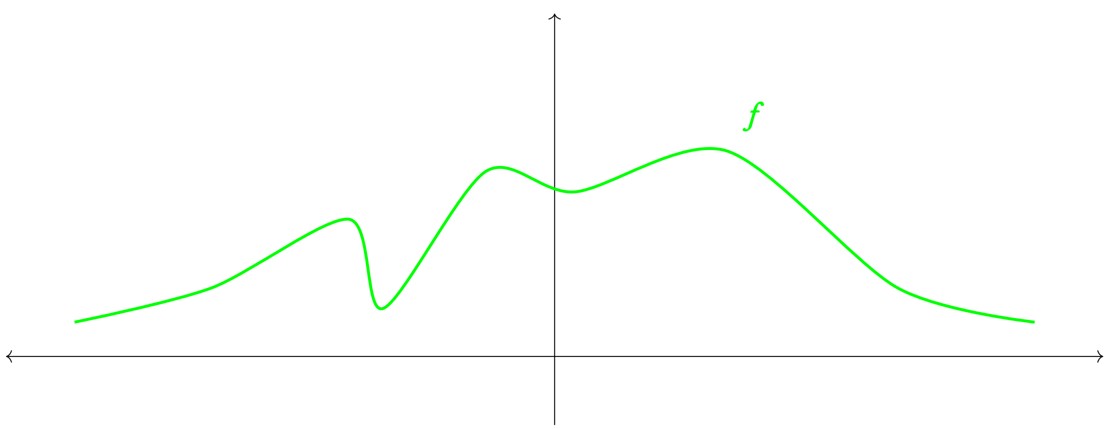

Figure 2.3: Function representing some density to illustrate the idea of distributions.

In experimental physics or engineering, a measurement device would be used to measure the density. In pure mathematics as is this context, the analogue of the measurement devices are the so called *Test Functions*.

Test Functions will be precisely defined later on with the rest of the theory presented. However, for now, the important think is to capture the idea of the roles these functions play. These functions should be well behaved in some sense, so properties like continuity and "localization" (in a sense that will be defined) are required.

In the context of the example given before of a function $f : \mathbb{R} \to \mathbb{R}$, the test functions will also have domain and codomain in $\mathbb{R}$, being functions of the form $\psi : \mathbb{R} \to \mathbb{R}$ that are zero outside a finite interval and be, for example, in the form of a hump inside this finite interval. This is illustrated in Figure 2.8 ahead.

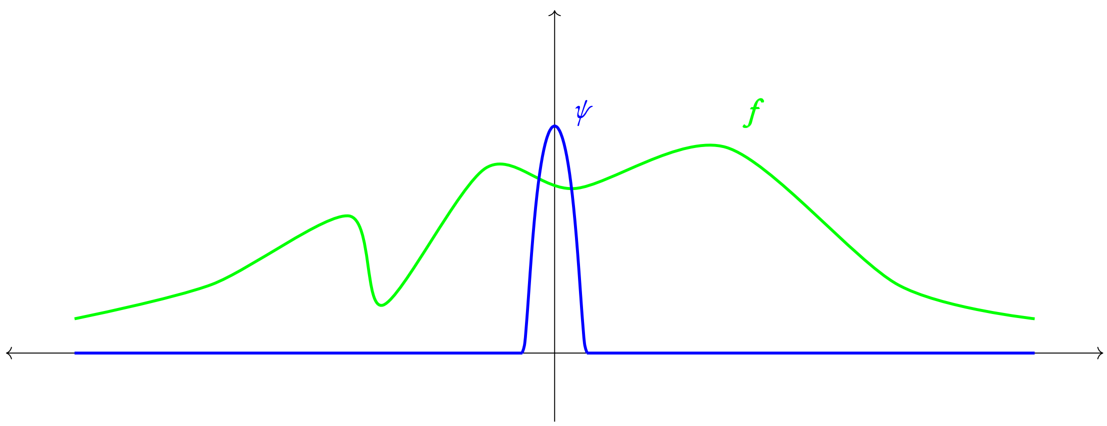

Figure 2.4: In green, a function f representing density and, in blue, a test function ψ. Their product $f\psi$ illustrates the idea of a "measurement" of the density at a specific localization.

The measurement itself will be made by taking the integral of the product of this two functions. In the real case discussed:

$$\int_{\mathbb{R}} f\psi \, d\lambda \in \overline{\mathbb{R}}.$$

The idea is to use many bumps to extract information from the function. In fact, given a function f, one could define a map

$$\psi \mapsto \int_{\mathbb{R}} f\psi \, d\lambda \in \overline{\mathbb{R}},$$

that takes test functions and associate to real numbers representing the value of the measurement.

The function f will be later on "substituted" and viewed as this map. This idea already resembles a lot of the concept of a functional on a dual space and the vision of a certain correspondence of a function and a functional, such as in the Riesz-Frechet Representation Theorem.

Some advantages in dealing with this map instead of trying to deal with the usual notion of a function can be viewed in the case of the Delta Function. In this case, the measurement will return:

$$\psi \mapsto \int_{\mathbb{R}} \delta\psi \, d\lambda = \psi(0),$$

which is the value of the test function at zero. Translations of the Delta could be considered to return the value of the test function at any point. This property of the Delta Function is known as the **Sampling Property**.

So, the central idea here is that functions are thought of as acting in the points of its domain and assigning values in its codomain. In distribution theory, there is a reinterpretation where entities like a function are thought of acting on test functions in a certain way.

2.2 Test Functions, Bump Functions And The Space Of Test Functions $\mathcal{D}(U)$

Let henceforth in this section M_n denote an n-dimensional manifold, $U \subseteq M_n$ an open subset and $\mathbb{K}$ the field $\mathbb{R}$ or $\mathbb{C}$. Here we firstly recall some basic notations:

Notation 1. *In this Section, the following notations will be fixed:*

1. $C(U) = C^0(U) := \{f : U \to \mathbb{K} \ : \ f \text{ is continuous}\}$. *The set* $C(M_n)$ *will simply be denoted by* C.
2. *For every* $k \in \mathbb{Z}_+$, $C^k(U) := \{f : U \to \mathbb{K} \ : \ f \text{ is } k\text{- times continuously differentiable}\}$. *The set* $C^k(M_n)$ *will simply be denoted by* C^k.

3. $C^{\infty}(U) := \bigcap_{k \in \mathbb{Z}_+} C^k(U) = \{f : U \to \mathbb{K} \,:\, f \text{ is } k\text{- times continuously differentiable for every } k \in \mathbb{Z}\}$. *This is the set of* ***smooth*** *functions on U. The set* $C^{\infty}(M_n)$ *will simply be denoted by* C^{∞}.

4. *A sub-index notation with a c symbol will be used to denote the restriction of some space to the functions of topological compact support. For example,* $C_c^k(U)$ *will denote the k- times continuously differentiable with compact support.*

The following definition will be very useful to make the notations a little bit shorter and some results more easier to state.

Definition 13. *A* ***multi-index*** *of size n is an element of* $(\mathbb{Z}_+)^n$*. The* ***length*** *of a multi-index* $\alpha = (\alpha_1, \ldots, \alpha_n) \in (\mathbb{Z}_+)^n$ *is the scalar* $|\alpha| := \alpha_1 + \cdots + \alpha_n$*. This definition can be use to make other definitions and notations more easily to deal with in the context of several variables. In particular, the following will be stated:*

1. $x^{\alpha} = x_1^{\alpha_1} \cdots x_n^{\alpha_n}$;
2. $\partial^{\alpha} = \dfrac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}}$.
3. $(Z_+)^n$ *can be endowed with a partial order were* $\alpha \geq \beta$ *if and only if* $\alpha_i \geq \beta_i$ *for all* $i \in \{1, \ldots, n\}$. *When* $\alpha \geq \beta$*, a multi-index binomial coefficient can be defined as*

$$\binom{\alpha}{\beta} := \binom{\alpha_1}{\beta_1} \cdots \binom{\alpha_n}{\beta_n}.$$

With multi-index notation, it is clear that $\varphi \in C^{\infty}(U)$ if and only if $\partial^{\alpha}\varphi \in C(U)$ for every multi-index α.

The considerations of the test functions in the previous section required some kind of continuity and "localization". With the previous definitions and notations it is possible to define in a precise manner the *Space of Test Functions*.

Definition 14. *Let $U \subseteq M_n$ be an open set. The **Space of Test Functions** on U is the set (indeed a $\mathbb{K}$-vector space) $\mathcal{D}(U) := C_c^\infty(U)$ of all arbitrarily often continuously differentiable $\psi : U \to K$ with compact support.*

The claim made in the previous definition is straightforward.

Proposition 4. *$\mathcal{D}(U)$ is a $\mathbb{K}$-vector space.*

Proof. The proof is trivial as M_n is a metric space so compact sets are closed and bounded. The group operations are defined pointwise in the obvious way so that, given $\varphi, \psi \in \mathcal{D}(U)$, $(\psi + \varphi)(x) := \psi(x) + \varphi(x)$ for all $x \in U$. In fact $(\psi + \varphi) \in \mathcal{D}(U)$ as the sum of functions of compact support has compact support and, since each one is arbitrarily differentiable, so its the sum. Scalar multiplication operation is treated the same way. □

Example 5. *The most famous type of test functions are the so called **Bump Functions** [†]. A **Circular Bump Function** in $\mathbb{R}^n$ is a function $\Psi : \mathbb{R}^n \to \mathbb{R}$ of the form:*

$$\Psi(x) = \exp\left(\frac{-1}{1 - \|x\|^2}\right) \cdot \mathbb{1}_{B_1} = \begin{cases} \exp\left(\frac{-1}{1 - \|x\|^2}\right), & \text{if } \|x\| < 1 \\ 0, & \text{otherwise.} \end{cases}$$

The graph of a circular bump function has the form shown in Figure 2.5 shown bellow for the $\mathbb{R}^2$ case.

† Some author use the term Bump Functions as a synonym to Test Functions

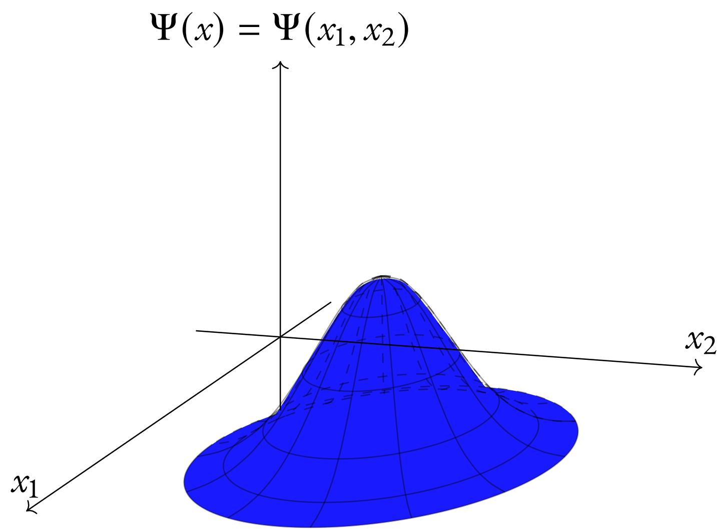

Figure 2.5: Example of a two dimensional bump function.

A ***Square Bump Function*** *in* $\mathbb{R}^n$ *is a function* $\Phi : \mathbb{R}^n \to \mathbb{R}$ *that is the product of one dimensional circular bump functions on* $\mathbb{R}$*, i.e., functions of the form:*

$$\Phi(x) = \Psi(x_1) \cdots \Psi(x_n).$$

2.3 Notion Of Convergence And Induced Topology on $\mathcal{D}(U)$

So, already being established that $\mathcal{D}(U)$ is a infinite dimensional vector space, the goal is to add a more structure to this space and for that matter a notion on convergence is very important. For that to be case, a metric, topology or norm should be defined.

In fact, every norm induces a metric and every metric induces a topology. However, latter on, it will be shown that defining just the a norm or a metric will not be sufficient to establish a specific convergence on a suitable topology.

The motivation for this concept of convergence is that the distributions should act continuously on the test functions and that the convergence to be defined gives the right notion of continuity.

Firstly, however, it is important to recall the possible notions of convergence to define the limit of sequences in vector spaces that are normed, metric or topological spaces.

Definition 15. *Let* $(V, \mathbb{F})$ *be a vector space over a field* $\mathbb{F}$*,* (v_j) *a sequence of elements in* V *and* $v \in V$*.*

1. *If* $\|\cdot\|$ *is a norm endowed in* $(V, \mathbb{F})$ *such that* $(V, \mathbb{F}, \|\cdot\|)$ *is a linear normed space, then, by definition:*
$$v_j \xrightarrow{\|\cdot\|} v \iff \left\|v_j - v\right\| \xrightarrow{n \longrightarrow \infty} 0.$$

2. *If* $d(\cdot,\cdot)$ *is a distance endowed in* $(V, \mathbb{F})$ *such that* $(V, \mathbb{F}, d)$ *is a linear normed space, then, by definition:*
$$v_j \xrightarrow{d} v \iff d(v, v_j) \xrightarrow{n \longrightarrow \infty} 0.$$

3. *If* τ *is a topology endowed in* $(V, \mathbb{F})$ *such that* $(V, \mathbb{F}, \tau)$ *is a topological vector space, then, by definition:*
$$v_j \xrightarrow{\tau} v \iff \forall U \in \tau \text{ with } v \in U, \ \exists N \in \mathbb{Z}_+ \text{ such that } v_j \in U \text{ for every } j \geq N.$$

It turns out that the metric and normed convergence are not sufficient and a more strong notion of convergence is needed. This convergence should require the two basic properties that characterize test functions, *i.e.*, the compact support and the arbitrarily differentiability.

For that matter, define a suitable topology (actually in the following the definition will be of sequential continuity, in the same way that will be done in the space of distributions, and this will be explained in an observation at the end of the definition of convergence for distributions):

Definition 16. *Given a sequence of test functions $(\varphi_j) \subseteq \mathcal{D}(U)$ and $\varphi \in \mathcal{D}(U)$, the* ***Canonical Distribution Convergence*** *(sometimes called* ***$\mathcal{D}$-convergence*** *for short) is defined by saying that $\varphi_j \xrightarrow{\mathcal{D}} \varphi$ if and only if*

1. *There exists a compact set $K \subseteq U$ such that $supp(\varphi_j) \subseteq K$ for all j;*

2. *For any multi-index $\alpha \in (\mathbb{Z}_+)^m$, $m \in \mathbb{Z}_+$, $\partial^\alpha \varphi_j$ converges to $\partial^\alpha \varphi$ uniformly.*

The first condition in Definition 16 is to avoid that the supports of the test functions in a sequence gets arbitrarily large so that the limit would not be a test function anymore. This is illustrated in Figure 2.9 bellow.

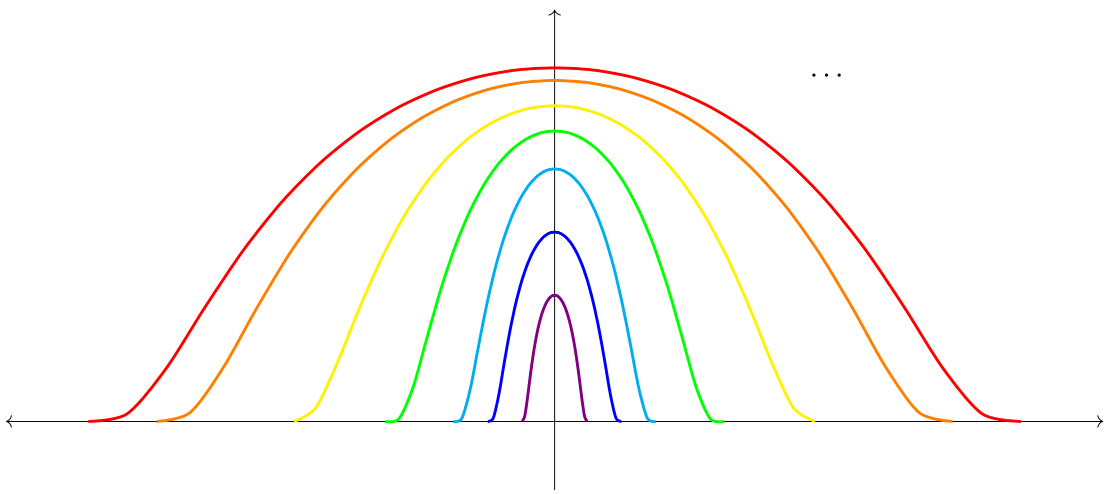

Figure 2.6: Illustration of the fact that the definition of $\mathcal{D}$ convergence requires that the support of the test function do not get arbitrarily large in analogy with light strips in a rainbow.

So, as it occurs in a rainbow with the light strips, it is wanted for test functions to have compact support so that a test function always tests bounded sets.

The second condition in Definition 16 is imposed to assure that every derivative gets arbitrarily close to the derivative of the limit function and sufficiently large index n in the sequence. This is illustrated in Figure 2.7 ahead.

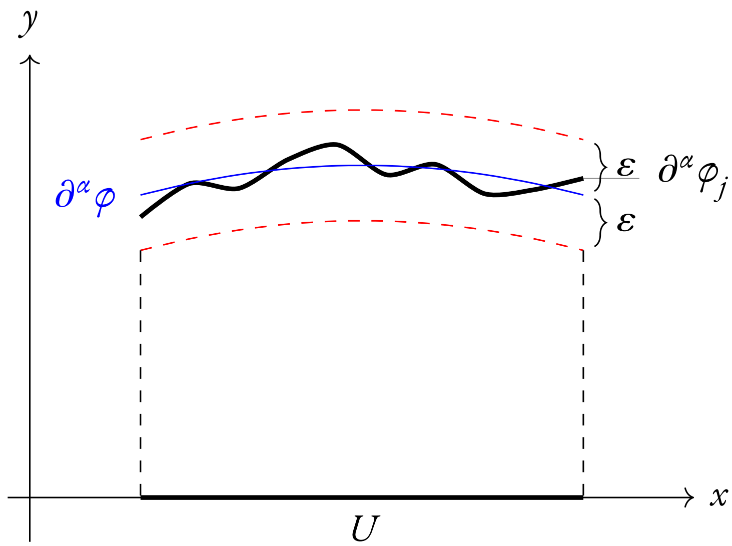

Figure 2.7: Illustration of the fact that the multi derivatives of the sequences of tests functions get arbitrarily close to the respect derivative of the limit test function.

In particular, for the 0 multi-index, the uniform convergence of the test functions themselves its assured.

It is a standard fact in analysis that uniform convergence is equivalent to convergence in the supremum norm. Therefore, the second requirement in Definition 16 can be rewritten as

$$\left\| \partial^\alpha \varphi_j - \partial^\alpha \varphi \right\|_\infty \xrightarrow{j \longrightarrow \infty} 0. \tag{2.1}$$

2.4 Distributions And The Space Of Distributions $\mathcal{D}'(U)$

The canonical distribution convergence endowed in $\mathcal{D}(U)$ in fact turns this space into a topological vector space, where the scalar multiplications, the group operation and the operation of taking the group inverses on the space are continuous map.

Now the concept of distributions and the *Space of Distributions* can be finally defined.

Definition 17. *A **Distribution** on $U \subseteq M_n$ is a continuous linear functional $T : \mathcal{D}(U) \to \mathbb{K}$, i.e., the following properties hold:*

1. *(Linearity):*
$$T(c_1\varphi_1 + c_2\varphi_2) = c_1 T(\varphi_1) + c_2 T(\varphi_2),$$
for all $c_1, c_2 \in \mathbb{K}$ and $\varphi_1, \varphi_2 \in \mathcal{D}(U)$;

2. *(Continuity):*
$$T(\varphi_j) \xrightarrow{j \longrightarrow \infty} T(\varphi) \text{ if } \varphi_i \xrightarrow{\mathcal{D}} \varphi.$$

*The set of all distributions on U is actually a vector space and it is called the **Space of Distributions** and it is denoted by $\mathcal{D}'(U)$. This vector space endowed with the weak topology ω forms a topological vector space and its the **Canonical Topological Vector Space of Distributions**.*

In requirement (2) of Definition 17, the notion of continuity is given by a notion of convergence. This is called **sequential continuity**. However, even though these notions are equivalent for metric spaces, they are not always equivalent in general. In fact, a topology is precisely defined in terms of convergence of **nets** (see [16]). Defining a topology is equivalent to specify all convergence nets. However, it can be show that, in this case, sequential continuity and continuity are equivalent since the space has local countable basis.

Example 6. *The **Dirac's Delta Distribution**, or more commonly, **Dirac's Delta Function** in $\mathbb{K}^n$ is a distribution $\delta \in \mathcal{D}'(\mathbb{K}^n)$ defined by*

$$\delta : \mathcal{D}(\mathbb{K}^n) \to \mathbb{K}$$
$$\varphi \mapsto \varphi(0).$$

Verifying the linearity is straightforward. The continuity its also not hard. In fact, let $\varphi \in \mathcal{D}(\mathbb{K}^n)$,

$(\varphi_j) \subseteq \mathcal{D}(\mathbb{K}^n)$ with $\varphi_j \xrightarrow{\mathcal{D}} \varphi$. Then, since the $\mathcal{D}$-convergence implies uniform (and therefore pointwise) convergence, $\varphi_j(x) \xrightarrow{j \longrightarrow \infty} \varphi(x)$ for every $x \in \mathbb{K}^n$. Taking $x = 0$, the continuity follows as $\delta(\varphi_j) = \varphi_j(0) \xrightarrow{j \longrightarrow \infty} \varphi(0) = \delta(\varphi)$.

Example 7. *Continuous Functions induce distributions and, therefore, sometimes are over viewed as distributions in some abuse of the term. In fact, let $f \in C(\mathbb{K}^n)$ be a continuous function. Then f induces a distribution T_f defined as*

$$\begin{aligned} T_f : \mathcal{D}(\mathbb{K}^n) &\longrightarrow \mathbb{K} \\ \varphi &\longmapsto T_f(\varphi) = \int_{\mathbb{K}^n} f\varphi \, d\lambda. \end{aligned}$$

The linearity comes from the fact that integral is a linear operator. The continuity comes similarly as what was done in Example 6, but here instead of the pointwise convergence, the more strong consequence of the $\mathcal{D}$-convergence, namely the uniform convergence, will be required:

$$\varphi_j \xrightarrow{\mathcal{D}} \varphi \implies \varphi_j \xrightarrow{unif} \varphi.$$

Then the claim is that $T_f(\varphi_j) \xrightarrow{j \longrightarrow \infty} T_f(\varphi)$. In fact, by the Dominated Convergence Theorem:

$$\lim_{j \longrightarrow \infty} T_f(\varphi_j) = \lim_{j \longrightarrow \infty} \int_{\mathbb{K}^n} \varphi_j \cdot f \, d\lambda = \int_{\mathbb{K}^n} \lim_{j \longrightarrow \infty} (\varphi_j \cdot f) \, d\lambda = \int_{\mathbb{K}^n} f \cdot \lim_{j \longrightarrow \infty} (\varphi_j) \, d\lambda = \int_{\mathbb{K}^n} f \cdot \varphi \, d\lambda = T_f(\varphi).$$

This types of distributions are usually called the ***Induced Distributions****. In fact, the same construction works for continuous functions defined on an open subset $U \subseteq \mathbb{K}^n$.*

Proposition 5. *Let $f : U \to \mathbb{K}$, $g : U \to \mathbb{K}$ be continuous functions and $T_f, T_g \in \mathcal{D}'(U)$ its induced distributions. Then:*

$$f = g \iff T_f = T_g.$$

Proof. The implication $f = g \implies T_f = T_g$ comes directly from the definition. Suppose now that $T_f = T_g$. This means that for every test function $\varphi \in \mathcal{D}(U)$:

$$\int_U f\varphi \, d\lambda = \int_U g\varphi \, d\lambda.$$

Since φ has compact support and f and g are continuous,

$$\int_U f\varphi \, d\lambda = \int_U g\varphi \, d\lambda < \infty.$$

Now, if

$$\int_E f d\lambda = \int_{M_n} f \cdot \mathbb{1}_E \, d\lambda = \int_E g \, d\lambda = \int_{M_n} g \cdot \mathbb{1}_E \, d\lambda$$

for every measurable set E, then $f = g$. However, every measurable function can be approximated by simple functions [†]. Every simple function in turn can be approximated by test functions. So take a sequence of test functions $(\psi)_j \longrightarrow \mathbb{1}_E$ pointwise. Then $(\psi)_j \cdot f \longrightarrow \mathbb{1}_E \cdot f$ and $(\psi)_j \cdot g \longrightarrow \mathbb{1}_E \cdot g$ pointwise. By the monotone convergence theorem:

$$\begin{cases} \lim\limits_{j \longrightarrow \infty} \int\limits_{M_n} \varphi_j \cdot f d\lambda = \int\limits_{M_n} \varphi_j \cdot f d\lambda = \int\limits_E f d\lambda, \\ \lim\limits_{j \longrightarrow \infty} \int\limits_{M_n} \varphi_j \cdot g \, d\lambda = \int\limits_{M_n} \varphi_j \cdot g \, d\lambda = \int\limits_E g \, d\lambda. \end{cases}$$

Hence, by the above equations, $f = g$ almost everywhere. However, since f and g are continuous, so is $f - g$. Thus, the set $A = \{x : f(x) \neq g(x)\}$ has measure zero. A in turn can be written as $A = [f - g]^{-1}((-\infty, 0) \cup (0, \infty))$, and, therefore is open. Being a open set of measure zero in the usual topology, $A = \emptyset$, which implies that $f = g$. □

[†] See [2].

Proposition 5 shows that indeed, continuous functions can be identified with distributions without any concept loss. Actually, there are some advantages in doing so, for example, obtaining tools like the Delta Function. In fact, more classes of functions can be identified with distributions:

Example 8. *Let $f: \mathbb{K}^n \to \mathbb{C}$ be **locally integrable**, i.e., for every compact set $K \subseteq \mathbb{K}^n$, $\int_K f \, d\lambda < \infty$. This is usually denoted by writing $f \in L^1_{Loc}(\mathbb{K}^n)$. Then f defines a distribution in the same fashion as done for continuous functions. In particular, changing the values of f on any set of measure zero will not change the distribution. For example, let*

$$x_+^s := \begin{cases} x^s, & \text{if } x > 0, \\ 0, & \text{if } x < 0, \end{cases}$$

defines a distribution if and only if $\mathfrak{Re}\{s\} > -1$.

There is another useful characterization of distributions which is presented in the following theorem:

Theorem 5. *A linear map $T \colon \mathcal{D}(U) \to \mathbb{K}$ is a distribution if and only if for all compact sets $K \subseteq U$ there exists a non negative integer $N \in \mathbb{Z}_+$ and a positive constant $C > 0$ such that for all tests functions $\varphi \in \mathcal{D}(U)$, if $supp(\varphi) \subseteq K$, then the value $T(\varphi)$ is bounded from above by $\left|T(\varphi)\right| \leq C \cdot \sum_{|\alpha| \leq N} \left\|\partial^\alpha \varphi\right\|_\infty$. In short notation:*

$$T \in \mathcal{D}'(U) \iff \underset{\substack{K \subseteq U, \\ K\, compact}}{\forall} \; \underset{N \in \mathbb{Z}_+}{\exists} \; \underset{C>0}{\exists} : \underset{\varphi \in \mathcal{D}(U)}{\forall}, \; supp(\varphi) \subseteq K \implies \left|T(\varphi)\right| \leq C \cdot \sum_{|\alpha| \leq N} \left\|\partial^\alpha \varphi\right\|_\infty .$$

Proof. ($\Longleftarrow$) : Suppose that the given estimate is valid.

Let $(\varphi_k) \subseteq \mathcal{D}(U)$, $\in \mathcal{D}(U)$ for all $k \in \mathbb{Z}_+$ and $\varphi_k \xrightarrow{\mathcal{D}} \varphi$. By hypothesis, there is a compact set $K \subseteq U$ with $\text{supp}(\varphi_k) \subseteq K$ for all k. Moreover, since the $\mathcal{D}$-convergence is valid, for all multi indices α, $\left\|D^\alpha \varphi_k - D^\alpha \varphi\right\|_\infty \xrightarrow{k \to \infty} 0$. It suffices to prove that $|T(\varphi_k) - T(\varphi)| \xrightarrow{k \to \infty} 0$. Indeed, by linearity and knowing that $\varphi - \varphi_k$ is again a test function, there exists a $N \in \mathbb{Z}_+$ and a $C > 0$ such that the estimate holds:

$$|T(\varphi_k) - T(\varphi)| = |T(\varphi_k - \varphi)| \leq C \cdot \sum_{|\alpha| \leq N} \left\| \partial^\alpha \varphi \right\|_\infty \xrightarrow{k \longrightarrow \infty} 0,$$

since the sum is finite over the multi indices. Hence, $T \in \mathcal{D}'(U)$.

$(\implies)$: Suppose that $T \in \mathcal{D}'(U)$.

The idea here is to prove by contraposition. Hence, the starting point is to negate the right hand side, which can be done by exchanging the quantifiers:

$$\underset{\substack{K \subseteq U, \\ K \text{ compact}}}{\exists} : \underset{N \in \mathbb{Z}_+}{\forall} \ \underset{C > 0}{\forall} \ \underset{\varphi \in \mathcal{D}(U)}{\exists}, \ \operatorname{supp}(\varphi) \subseteq K \wedge \left| T(\varphi) \right| > C \cdot \sum_{|\alpha| \leq N} \left\| \partial^\alpha \varphi \right\|_\infty,$$

where here $\wedge$ denotes the logical operator "and". The goal is to prove that T can not be continuous. The above expression means that there exist a compact set such that no matter what constants N and C are chosen, there is always a corresponding φ. Choose $N = C = k \in \mathbb{Z}_+$. Therefore:

$$|T(\varphi_k)| > k \cdot \sum_{|\alpha| \leq k} \left\| \partial^\alpha \varphi \right\|_\infty.$$

The zero multi index is as well considered in the sum, which means that the supremum norm of each φ_k is also considered. The idea is to define new test functions ψ_k that will be given by a rescaling of the φ_k functions. Since the absolute value $|T(\varphi_k)|$ is increasing with k and increases faster than the supremum norm of φ_k, take

$$\psi_k := \frac{1}{|T(\varphi_k)|} \varphi_k.$$

By the definition of the ψ_k's and the consideration about the supremum norm, the ψ_k's converges uniformly to the zero function when k goes to infinity. The same argument works for every multi index, so that every derivative of ψ_k behaves the same way. This combined with the fact that all φ_k (and therefore all ψ_k) has support in K, assures the $\mathcal{D}$-convergence. Hence $\psi_k \xrightarrow{\mathcal{D}} 0$. However, the images under T don't converge to zero since:

$$|T(\psi_k)| = \frac{1}{|T(\varphi_k)|}|T(\varphi_k)| = 1 \forall k \implies |T(\psi_k)| \overset{k\longrightarrow\infty}{\not\longrightarrow} 0.$$

Therefore, the map T is not continuous and the proof is completed. □

This means that whenever this estimate is valid, the continuity is assured and therefore, a distribution. Indeed, some authors use this estimate given in Theorem 5 as the definition of a distribution. This motivates the following definition:

Definition 18. *Let $U \subseteq M_n$ be an open set. Then*

$$T \in \mathcal{D}'(U) \iff \underset{\substack{K\subseteq U,\\ K\,compact}}{\forall}\ \underset{N\in\mathbb{Z}_+}{\exists}\ \underset{C>0}{\exists} : \underset{\varphi\in\mathcal{D}(U)}{\forall},\ supp(\varphi) \subseteq K \implies \left|T(\varphi)\right| \leq C \cdot \sum_{|\alpha|\leq N} \left\|\partial^\alpha \varphi\right\|_\infty.$$

*If the constant N does not depend on the compact set K, the distribution is said to have **order N**, or be a **N-order distribution**. This distributions are denoted as $\mathcal{D}'_N(U)$. When one wants to refer to the specific value of N, one could say the distribution has **exactly** order N. The set of distributions of finite order is often denoted as $\mathcal{D}'_F(U)$.*

Example 9. *Every locally integrable function can be viewed as a 0-order distribution, i.e., $f \in L^1_{Loc}(U) \implies T_f \in \mathcal{D}'_0(U)$.*

Example 10. *Let μ be any positive Radon measure. Then μ defines a distribution T_μ by*

$$T_\mu(\varphi) := \int \varphi\, d\mu.$$

In fact, the linearity is obvious. Moreover

$$|T_\mu(\varphi)| \leq \mu(supp(\varphi))\|\varphi\|_\infty,$$

and hence, Theorem 5 can be immediately applied, showing that T_μ is indeed a distribution.

Example 11. *Any measure on M_n can be viewed as a 0-order distribution.*

Example 12. *If T is a distribution defined as*

$$T(\varphi) = \partial^\alpha \varphi(x_0),$$

for every test function $\varphi \in \mathcal{D}(U)$ and a specific point $x_0 \in M_n$, then φ is a $|\alpha|$-order distribution.

Example 13. *Let (x_j) be a sequence of points in M_n without a limit point. If T is a distribution defined as*

$$T(\varphi) = \sum_j \partial^{\alpha_j} \varphi(x_j),$$

then $T \in \mathcal{D}'_F(U)$ if and only if $\sup(|\alpha_j|) < \infty$. In case $T \in \mathcal{D}'_F(U)$, the exact order of T is $\sup(|\alpha_j|)$.

2.5 Regular Distributions

There is a problem concerning distributions that is analogous to the problem presented in the formulation of Fourier Series in general Hilbert Spaces. That is the fact that a distribution is not a function in the usual sense and therefore, it is not possible to evaluate a distribution $T \in \mathcal{D}'(M_n)$ in some point $x \in M_n$ of the space.

Nonetheless, it is still meaningful to talk about "evaluating a distribution on an open set collectively". A distribution $T \in \mathcal{D}'(M_n)$ is said to **vanish** in an open set $U \subseteq M_n$ if $T(\varphi) = 0$ for all test functions φ such that $\mathrm{supp}(\varphi) \subseteq U$. Likewise, a distribution $T \in \mathcal{D}'(M_n)$ is said to **agree** with a function f on a open set $U \subseteq M_n$ if $T(\varphi) = \int f \cdot \varphi \, d\lambda$ for all φ with $\mathrm{supp}(\varphi) \subseteq U$. This is actually one of the reasons one uses the name "test functions", because it is possible to image that functions $\varphi \in \mathcal{D}(M_n)$ are used to "test" or "detect" the values of a function on a open set. However, as presented in the previous section, locally integrable functions induce distributions and those distributions induce functions that are equal almost everywhere. This are some special distributions that are called *Regular Distributions*.

Recall that a function $f : \mathbb{K}^n \longrightarrow \mathbb{C}$ is locally integrable if for every compact set $K \subseteq \mathbb{K}^n$, $\int_K f \, d\lambda < \infty$. For this functions, the notation $f \in L^1_{\mathrm{Loc}}(\mathbb{K}^n)$. It is obvious that every integrable function is locally integrable but the converse does not always hold. Likewise, every continuous function is locally integrable, but not every locally integrable function is continuous.

With the last result in the previous section given by Theorem 5, it is possible to prove formally that in fact locally integrable functions correspond uniquely, except maybe in a set of measure zero, in a one to one relation with an induced distribution T_f. However, the proof is technically more complicated compared to the case where the function was continuous. That is due to the fect that, when the function considered were continuous, if it is nonzero at some point, there would exist an open set with nonzero measure where the function would be nonzero. This is not the case for general locally integrable functions.

In order to prove this result an important result from measure theory will be required.

Theorem 6 (Lebesgue Differentiation Theorem)**.** *Suppose that $f \in L^1_{Loc}(\mathbb{K}^n)$. Then, λ-almost all x:*

$$\lim_{r \longrightarrow 0} \frac{1}{\lambda(B_r(x))} \int\limits_{B_r(x)} |f(y) - f(x)| \, dy = 0,$$

and

$$\lim_{r \longrightarrow 0} \frac{1}{\lambda(B_r(x))} \int\limits_{B_r(x)} f(y) dy = f(x).$$

Proof. See [8]. □

Proposition 6. *Let $f : M_n \to \mathbb{K}$ be a locally integrable function. Then T_f induces a distribution T_f that is uniquely related to f except maybe in a set of zero measure i.e.,*

$$f = g \; a.e. \iff T_f = T_g.$$

Proof. Without loss of generality, consider $M_n = \mathbb{K}^n$. It then suffices to show that the estimate of Theorem 5 is valid. So let $f \in L^1_{\text{Loc}}(\mathbb{K}^n)$ and define as before:

$$\begin{aligned} T_f : \mathcal{D}(\mathbb{K}^n) &\longrightarrow \mathbb{K} \\ \varphi \longmapsto T_f(\varphi) &= \int_{\mathbb{K}^n} f\varphi \, d\lambda. \end{aligned}$$

By definition then:

$$|T_f(\varphi)| \leq \int_{\mathbb{K}^n} |f| \cdot |\varphi| \, d\lambda = \int_{\text{supp}(\varphi)} |f| \cdot |\varphi| \, d\lambda \leq \int_{\text{supp}(\varphi)} |f| \, d\lambda \cdot \|\varphi\|_\infty = \int_K |f| \, d\lambda \cdot \|\varphi\|_\infty,$$

for every compact set $K \supseteq \text{supp}(\varphi)$. Therefore, it is straight forward that

$$|T(\varphi)| \leq C \cdot \sum_{|\alpha|<m} \|D^\alpha \varphi\|_\infty,$$

with $C = \int_K |f| \, d\lambda$ and $m = 0$.

Now, in order to prove that if $T_f = T_g$ then $f = g$ almost everywhere for two locally integrable functions f and g, first notice that $T_f = T_g$ means that

$$\int_{M_n} f\varphi \, d\lambda = \int_{M_n} g\varphi \, d\lambda$$

for every test function $\varphi \in \mathcal{D}(M_n)$, which implies that:

$$\int_{M_n} (f-g)\varphi \, d\lambda = 0,$$

so it suffices to prove that if $T_f = 0$, then $f = 0$ λ-almost everywhere.

Now, notice that one can write for every density point x:

$$f(x) = \frac{1}{\lambda(B_r(x))} \int_{B_r(x)} \varphi(t)\,[f(x) - f(t)]\, dt,$$

where φ is a bump function that is equal to 1 at the ball of radius r and 0 outside the ball of radius $2r$ and the integration is with respect to the Lebesgue measure. Indeed, the above equality is straightforward to very and it is just a simple calculation. However, due to Theorem 6, one has that

$$\frac{1}{\lambda(B_r(x))} \int_{B_r(x)} \varphi(t)\,[f(x) - f(t)]\, dt \xrightarrow{r \longrightarrow 0} 0,$$

so $f = 0$ λ-almost everywhere and the result is proven. □

Definition 19. *A distribution $T \in \mathcal{D}'(U)$ is called **regular** if there is a locally integrable function f such that $T = T_f$. A distribution that is not regular is called a **singular** distribution.*

These distributions in some sense behave like normal functions as discussed before. Of course not every distribution is regular. In fact the first distribution presented, the Delta, is not regular.

Example 14. *The Delta distribution is not regular, i.e., there is no locally integrable function that induces δ. Consider here $M_n = \mathbb{K}^n$ and functions $f : \mathbb{K}^n \to \mathbb{K}$. Then $\delta \neq T_f$ for every locally integrable function.*

To see this, suppose such f exists, i.e., $f \in L^1_{Loc}(K^n)$ with $\delta(\varphi) = T_f(\varphi)$ for all $\varphi \in \mathcal{D}(K^n)$. Then

$$\varphi(0) = \delta(\varphi) = T_f(\varphi) = \int_{K^n} f \cdot \varphi \, d\lambda.$$

Since $f \in L^1_{Loc}$, it is absolutely integrable on any compact set $K \subseteq \mathbb{K}^n$, in particular in the unit ball B_1. The unit ball is in turn the union of countable circular rings R_k delimited by the circumferences of radius $r_k = 1/k$, except for the point at the origin O. Hence

$$\int_{B_1} |f| \, d\lambda = \int_{B_1 \setminus O} |f| \, d\lambda = \int_{\bigcup R_k} |f| \, d\lambda = a < \infty,$$

for some constant a. However, since this union is disjoint, the Monotone Convergence Theorem assures that

$$\int_{\bigcup R_k} |f| \, d\lambda = \sum_{k=1}^{\infty} \int_{R_k} |f| \, d\lambda = a < \infty.$$

Therefore, the general term goes to zero since the series converges. So there must exist some constant $k_0 \in \mathbb{Z}_+$ such that

$$\sum_{k=k_0}^{\infty} \int_{R_k} |f| \, d\lambda = b < 1.$$

Then, there must exist $\varepsilon > 0$ such that, denoting the ε-ball $\|x\| \leq \varepsilon$ by B_ε,

$$\int_{B_\varepsilon} |f| \, d\lambda = b < 1.$$

The idea to finish the proof is to take test functions that are concentrated at zero, since δ is also "concentrated" at zero. Take then the Bump Function:

$$\Psi_\varepsilon(x) = \exp\left(\frac{-1}{1-\left(\frac{\|x\|^2}{\varepsilon}\right)}\right)\cdot \mathbb{1}_{B_\varepsilon} = \begin{cases}\exp\left(\frac{-1}{1-\left(\frac{\|x\|^2}{\varepsilon}\right)}\right), & \text{if } \|x\| < \varepsilon\\ 0, & \text{otherwise.}\end{cases}$$

Then,by the definition of the δ and Ψ_ε and still supposing the regularity of f:

$$\delta(\Psi_\varepsilon) = \Psi_\varepsilon(0) = \int_{\mathbb{K}^n} f\cdot \Psi_\varepsilon\, d\lambda,$$

and since f and Ψ_ε are both nonnegative:

$$\begin{aligned}\Psi_\varepsilon(0) = \int_{\mathbb{K}^n} f\cdot \Psi_\varepsilon\, d\lambda = \int_{B_\varepsilon} f\cdot \Psi_\varepsilon\, d\lambda = \left|\int_{B_\varepsilon} f\cdot \Psi_\varepsilon\, d\lambda\right| \leq \int_{B_\varepsilon} |f|\cdot|\Psi_\varepsilon|\, d\lambda \leq \|\Psi_\varepsilon\|_\infty \cdot \int_{B_\varepsilon} |f|\, d\lambda\\ < \|\Psi_\varepsilon\|_\infty\cdot b < \|\Psi_\varepsilon\|_\infty,\end{aligned}$$

clearly a contradiction, since $\|\Psi_\varepsilon\|_\infty$ can not be less than $\Psi_\varepsilon(0)$. Thus, the Delta distribution is not regular.

2.6 Compatibility Between $\mathcal{D}(U)$ and $\mathcal{D}'(U)$ And The Duality Pairing

It is already established in the text that both $\mathcal{D}'(U)$ and $L^1_{\text{Loc}}(U)$ are vector spaces and, in fact, $\mathcal{D}'(U)$ is actually a topological vector space. Moreover, two different continuous functions give rise to two different distributions on $\mathcal{D}'(U)$ and the same property still holds for every function on $L^1_{\text{Loc}}(U)$ up to sets of measure zero. Therefore, there is no loss of information in the process of going from $L^1_{\text{Loc}}(U)$ and in fact, $\mathcal{D}'(U)$ can be viewed as a generalisation of $L^1_{\text{Loc}}(U)$ since it also includes singular distributions like the Delta. The goal in this section is then to ask the natural question that arises immediately from this extension, *i.e.*, if the operations in these two vector spaces are consistent with each other.

The linearity is straight forwarded since, for $\alpha \in \mathbb{K}$ and $f, g \in L^1_{\text{Loc}}(M_n)$:

$$[T_{f+\alpha g}](\varphi) = \int_{M_n} (f+\alpha g) \cdot \varphi \, d\lambda = \int_{M_n} f \cdot \varphi \, d\lambda + \int_{M_n} \alpha g \cdot \varphi \, d\lambda = \int_{M_n} f \cdot \varphi \, d\lambda + \alpha \int_{M_n} g \cdot \varphi \, d\lambda$$
$$= T_f(\varphi) + T_g(\varphi) = [T_f + \alpha T_g](\varphi),$$

for every $\varphi \in \mathcal{D}(U)$.

So it is clear that the operations are consistent between $L^1_{\text{Loc}}(U)$ and the subspace of regular distributions on $\mathcal{D}'(U)$. The natural question that arises is that if this compatibility is still valid for all $\mathcal{D}(U)$ and $\mathcal{D}'(U)$. It turns out that indeed this makes sense because actually the distributions induced by elements in $\mathcal{D}(U)$ are dense in the space of all distributions $\mathcal{D}'(U)$. This will be proved at the end of the Chapter using a tool called *Approximations To The Identity* or *Mollifiers*.

Therefore, it is usual to introduce the following definition.

Definition 20. *Given a test function $\varphi \in \mathcal{D}(U)$ and a distribution $T \in \mathcal{D}(U)$, the evaluation of a distribution can be represented as*

$$\langle T, \varphi \rangle := T(\varphi).$$

*This is called the **Duality Pairing** notation between $\mathcal{D}(U)$ and $\mathcal{D}'(U)$.*

This may appear useless at first but will be very practical, especially dealing with derivatives and other operations. This clearly emphasizes a bilinear characterization:

$$\langle \cdot, \cdot \rangle : \mathcal{D}'(U) \times \mathcal{D}(U) \longrightarrow \mathbb{K}.$$

It is important to emphasize that, although the bracket notation here really reminds the usual inner product in function spaces given by integration, this only is a notation when dealing with all distributions since not all distributions are regular, and thus, it does not make sense to integrate their product with a test function in an obvious way.

2.7 Multiplication of Smooth Functions And Test Functions

The operations of scaling and sum of distributions are pretty clear. However, the question about multiplying two distributions is quite more delicate. In fact, the multiplication $T \cdot J$ of two distributions $T, J \in \mathcal{D}'(U)$ is not obvious at all to define. Indeed, such a general multiplication for general distributions would lose important properties of the usual function multiplication and, therefore, lose its meaning as an extension of the usual multiplication.

Firstly, one could ask a more basic question of what could be the multiplication of a distribution by a smooth function. So let $T \in \mathcal{D}'(U)$ and $f \in C^{\infty}(U)$. The goal is to define the product $T \cdot T_f$. Consider in the first place the case where T is a regular distribution. In this case, $T = T_g$, for some $g \in L^1_{\text{Loc}}(U)$. In order to the operations to be compatible, one could wish to define $T_f \cdot T_g(\varphi) = T_{f \cdot g}(\varphi)$ for any test function $\varphi \in \mathcal{D}(U)$. Indeed, doing this:

$$[T_f + T_g](\varphi) := T_{f \cdot g}(\varphi) = \int\limits_U (f \cdot g)\varphi \, d\lambda = \int\limits_U g(f \cdot \varphi) \, d\lambda,$$

and since $f \cdot \varphi$ is again a test function (indeed, both f and φ are smooth and φ has compact support, making the product a smooth function with compact support):

$$[T_f + T_g](\varphi) := T_{f \cdot g}(\varphi) = T_g(f \cdot \varphi).$$

In fact, the general definition also makes sense to be stated for distributions T that are not regular, as long as T_f is still defined the same way.

Definition 21. *Let $T \in \mathcal{D}'(U)$ and $f \in C^{\infty}(U)$. The product of the distributions T_f and T is denoted* *by $T \cdot T_f$ (or sometimes $f \cdot T$ in an abuse of language identifying functions with distributions) and is defined as:*

$$\langle f \cdot T, \varphi \rangle := \langle T, f \cdot \varphi \rangle,$$

for all test functions $\varphi \in \mathcal{D}(U)$.

However, it should be proved that this object still fulfills the two properties of distributions, *i.e.*, linearity and continuity.

Proof. The linearity is straightforward. Indeed, given $\varphi, \psi \in \mathcal{D}(U)$ and $\alpha \in \mathbb{K}$:

$$\begin{aligned}\langle T \cdot T_f, \varphi + \alpha\psi \rangle = \langle T, f \cdot (\varphi + \alpha\psi) \rangle = \langle T, f\varphi + \alpha f\psi \rangle = T(f\varphi + \alpha f\psi) = T(f\varphi) + \alpha T(f\psi) \\ = \langle T, f\varphi \rangle + \alpha \langle T, f\psi \rangle = \langle T \cdot T_f, \varphi \rangle + \alpha \langle T \cdot T_f, \psi \rangle.\end{aligned}$$

To prove the continuity, Theorem 5 will be used. By the Leibniz Rule For Differentiation (in multi-index notation):

$$D^\alpha (f \cdot \varphi) = \sum_{\beta \leq \alpha} \binom{\alpha}{\beta} D^\beta f \cdot D^{\alpha - \beta} \varphi.$$

So, by Theorem 5,

$$T \in \mathcal{D}'(U) \iff \underset{\substack{K \subseteq U, \\ K \text{compact}}}{\forall} \ \underset{N \in \mathbb{Z}_+}{\exists} \ \underset{C > 0}{\exists} \ : \underset{\tilde{\varphi} \in \mathcal{D}(U)}{\forall}, \ \operatorname{supp}(\tilde{\varphi}) \subseteq K \implies \left| T(\tilde{\varphi}) \right| \leq C \cdot \sum_{|\alpha| \leq N} \left\| \partial^\alpha \tilde{\varphi} \right\|_\infty.$$

It is already known that T is a distribution. Hence, the goal is to show that an analogous inequality holds for $T \cdot T_f$. Indeed, by the definition of the product of the distributions and considering the estimate given that T is a distribution:

$$|T \cdot T_f(\varphi)| = |T(f \cdot \varphi)| \leq C \cdot \sum_{|\alpha| \leq N} \left\|\partial^\alpha (f \cdot \varphi)\right\|_\infty = C \cdot \sum_{|\alpha| \leq N} \left\| \sum_{\beta \leq \alpha} \binom{\alpha}{\beta} D^\beta f \cdot D^{\alpha-\beta} \varphi \right\|_\infty$$
$$\leq C \cdot \sum_{|\alpha| \leq N} \sum_{\beta \leq \alpha} \binom{\alpha}{\beta} \left\|D^\beta f \cdot D^{\alpha-\beta}\varphi\right\|_\infty \leq C \cdot \sum_{|\alpha| \leq N} \sum_{\beta \leq \alpha} \binom{\alpha}{\beta} \left\|D^\beta f\right\|_\infty \cdot \left\|D^{\alpha-\beta}\varphi\right\|_\infty.$$

Well, since α is a finite multi-index, the sum that runs over β is a sum of finitely many terms. Moreover, the term $\left\|D^\beta f\right\|_\infty$ is finite and does not depend on the test function φ. Also, summing along the indices $\alpha - \beta$ is accounted for in the sum over $|\alpha| < N$. Hence, the sum over β that runs across all combinations $\binom{\alpha}{\beta}$ can be put in a constant term whose product with the constant C results in a new constant $\tilde{C}$ leading to

$$|T \cdot T_f(\varphi)| \leq \tilde{C} \sum_{|\alpha| \leq N} \left\|D^\alpha \varphi\right\|_\infty,$$

proving that $T \cdot T_f$ is indeed a distribution. Therefore, the product of a distribution and a distribution induced by a smooth function is well defined and again a distribution. □

2.8 Coordinate Transformations and Pullbacks

It is also important in the theory to understand how a change of coordinates works in the context of distributions.

As a starting point, consider $M_n = \mathbb{K}^n$ and a linear map $A : \mathbb{K}^n \to \mathbb{K}^n$. If this map is invertible, it can be seen as a change of coordinates.

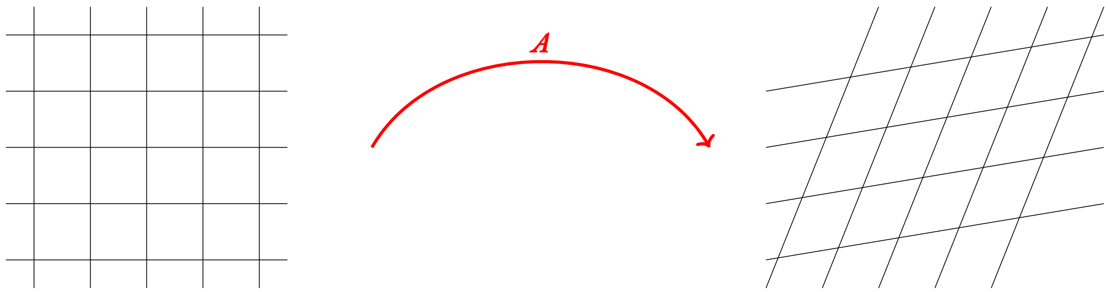

Figure 2.8: Change of coordinate by invertible linear map.

The invertible linear A can stretch, rotate or reflex the original grid that represents the original basis. However, many properties does not change with this coordinate transformation. Here, one of the most important ones is that a locally integrable function is still locally integrable with respect to this new basis. More precisely

$$f \in L^1_{\text{Loc}}(\mathbb{K}^n) \implies f \circ A \in L^1_{\text{Loc}}(\mathbb{K}^n).$$

The central point here is then to understand how this change of coordinates connects the two corresponding distributions (induced by f and $f \circ A$). However, by definition for $f \in L^1_{\text{Loc}}(\mathbb{K}^n)$ and $\varphi \in \mathcal{D}(\mathbb{K}^n)$

$$\langle T_{f \circ A}, \varphi \rangle = \int\limits_{\mathbb{K}^n} (f \circ A) \cdot \varphi \, d\lambda,$$

and, doing a change of variables (since A is invertible):

$$\begin{aligned}\langle T_{f \circ A}, \varphi \rangle &= \frac{1}{|\det(A)|} \int_{\mathbb{K}^n} f(\underbrace{A(x)}_{y}) \cdot \varphi(x) \cdot \underbrace{|\det(A)|\, dx}_{dy} = \frac{1}{|\det(A)|} \int_{\mathbb{K}^n} f(y) \varphi(A^{-1} y)\, dy \\ &= \left\langle T_f, \frac{1}{|\det(A)|} \varphi \circ A^{-1} \right\rangle.\end{aligned}$$

This calculation motivates the definition for non-regular distributions to be the same result.

So let $T \in \mathcal{D}(\mathbb{K}^n)$ and $A : \mathbb{K}^n \longrightarrow \mathbb{K}^n$ be an invertible linear map. † Then denote by $T \circ A$ the distribution induced by the change of coordinates A defined by

$$\langle T \circ A, \varphi \rangle := \left\langle T, \frac{1}{|\det(A)|} \varphi \circ A^{-1} \right\rangle.$$

A translation may also be considered. Let A be an invertible linear map as before and $b \in \mathbb{K}^n$ and consider the new map A_b defined by

$$\begin{aligned} A_b : \mathbb{K}^n &\longrightarrow \mathbb{K}^n \\ x &\longmapsto A(x) + b. \end{aligned}$$

The inverse will be

$$\begin{aligned} A_b^{-1} : \mathbb{K}^n &\longrightarrow \mathbb{K}^n \\ y &\longmapsto A^{-1}(y - b). \end{aligned}$$

Therefore, one could also define $T \circ A_b$ the distribution induced by the change of coordinates A_b defined by

$$\langle T \circ A, \varphi \rangle := \left\langle T, \frac{1}{|\det(A)|} \varphi \circ A_b^{-1} \right\rangle.$$

In more generality, consider the following definition:

Definition 22. *Let A, B, C be sets and consider two maps $f : B \to C$ and $g : A \to B$. The* ***pullback*** *of the map f induced by g is the map $g := f \circ g$.*

Let $U_1, U_2 \subseteq \mathbb{R}^n$ be open sets and $F : U_1 \to U_2$ be a C^∞ diffeomorphism. If $f \in C^0(U_2)$ is a

†Here this will not be put in a definition in the usual way because at the end a more general definition will be made considering more general coordinate transformations.

complex function, then $F = f \circ F : U_1 \longrightarrow \mathbb{C}$ is in $C^0(U_1)$ and thus can be viewed as an element of $\mathcal{D}'(U_1)$ Then, its action in the change of coordinates induced by the diffeomorphism will be

$$\begin{aligned}\langle F\star f, \varphi\rangle &= \int_{U_1} f(F(x))\varphi(x)\,dx = \int_{U_2} f(y)\varphi(F^{-1}(y))|\det\big(D(F^{-1})(y)\big)|\,dy \\ &= \langle f, \varphi\circ(F^{-1})\cdot|\det\big(D(F^{-1})\big)|\rangle = \langle f, (F^{-1})\star\varphi\cdot|\det\big(D(F^{-1})\big)|\rangle.\end{aligned}$$

Therefore, the following definition can be formalized

Definition 23. *Let $U_1, U_2 \subseteq \mathbb{R}^n$ be open sets and $F : U_1 \longrightarrow U_2$ be a C^∞ diffeomorphism. Then, for any $T \in \mathcal{D}'(U_2)$, the* ***pullback distribution*** *of T under F is defined as:*

$$\langle F\star T, \varphi\rangle := \langle T, (F^{-1})\star\varphi\cdot|\det\big(D(F^{-1})\big)|\rangle,$$

for all $\varphi \in \mathcal{D}(U_1)$.

It can be showed that indeed that actually $F\star T \in \mathcal{D}'(U_1)$ and that the map $\mathcal{D}'(U_2) \longrightarrow \mathcal{D}'(U_1)$, $T \longmapsto F\star T$ is linear and sequentially continuous. If fact, this theory can be done in more generality for a **submersion**[‡] F. For more details, see [11].

Observation 2. *As a final remark in this section, it is worth pointing out that the general idea for extending new concepts like this to the theory of distributions, in general, have the same central idea. This works in general for operations well defined for the test functions $\varphi \in \mathcal{D}(U)$ (e.g. derivatives, Fourier Transforms, convolutions, etc). More precisely, let L be an operator such that $L(f) = Lf$ is well defined for some class of locally integrable functions f (and thus a natural concept arises for the associated regular distributions) and*

$$\langle Lf, \varphi\rangle = \langle f, L^*\varphi\rangle,$$

[‡] $F \in C^\infty(U_1, U_2)$ is a submersion if DF is surjective at each point of U_1.

for some operation L^*. Then, given an arbitrary distribution $T \in \mathcal{D}'(U)$, one could define LT in the same way:

$$\langle LT, \varphi \rangle = \langle T, L^* \varphi \rangle.$$

2.9 Derivatives Of Distributions

One of the big advantages in this extension, *i.e.*, working with distributions, is that a construction of a differentiation allow them to always be differentiated. Moreover, this definition extends the usual notion of derivatives of actual functions, another reason why distributions are very often called generalized functions.

The general idea to come up with the definition is the following. Consider $f : \mathbb{K}^n \to \mathbb{C}$, $f \in C^1(\mathbb{K}^n)$ and $T_f \in \mathcal{D}'(\mathbb{K}^n)$ the regular distribution induced by f. Then, the partial derivatives $\partial_i f$, $i \in \{1, \ldots, n\}$, of f are well defined and are continuous functions from $\mathbb{K}^n$ to $\mathbb{C}$. Hence, the partial derivatives induce regular distributions $T_{\partial_i f}$. This distributions induced by the derivatives is what one would expect to be defined as the derivatives of the distribution. In fact, for every $\varphi \in \mathcal{D}(U)$:

$$\langle T_{\partial_i f}, \varphi \rangle = \int_{\mathbb{K}^n} \partial_i f \varphi \, d\lambda.$$

The idea here is to use integration by parts. In fact, given an open bounded subset $\Omega \subseteq \mathbb{K}^n$ with a piecewise smooth boundary $\partial\Omega$, then

$$\int_{\Omega} \partial_i f \varphi \, d\lambda = \int_{\partial\Omega} f \varphi \, d\lambda - \int_{\Omega} f \partial_i \varphi \, d\lambda.$$

However, φ is compactly supported. Therefore, one could choose an open set $\Omega \supseteq \mathrm{supp}(\varphi)$ such that the integral outside Ω is zero. In fact, integrating in all $\mathbb{K}^n$ (one could think sending the boundary $\partial\Omega$

to infinity):

$$\int_{\mathbb{K}^n} \partial_i f \varphi \, d\lambda = -\int_{\mathbb{K}^n} f \partial_i \varphi \, d\lambda.$$

Therefore, the natural definition would be

$$\langle T_{\partial_i f}, \varphi \rangle := -\langle T_f, \partial_i \varphi \rangle = -\int_{\mathbb{K}^n} f \partial_i \varphi \, d\lambda = T_f(\partial_i \varphi).$$

The idea is to take this result as the definition for a general distribution $T \in \mathcal{D}'(U)$. The general idea in the observation in the final of the last section can be applied here, considering $L = \partial_i$ and the "adjoint" $L^* = -\partial_i$.

Definition 24. *Let $T \in \mathcal{D}'(\mathbb{K}^n)$, then the **$i^{th}$ partial derivative distribution**, denoted by $\partial_i T$ is defined by*

$$\langle \partial_i T, \varphi \rangle := -\langle T, \partial_i \varphi \rangle,$$

for all test functions $\varphi \in \mathcal{D}(\mathbb{K}^n)$. Hence, for every multi-index $\alpha \in (\mathbb{Z}_+)^k$, the partial derivative distribution with respect to α can be defined as

$$\langle \partial^\alpha T, \varphi \rangle = (-1)^{|\alpha|} \langle T, \partial^\alpha \varphi \rangle,$$

for all test functions $\varphi \in \mathcal{D}(\mathbb{K}^n)$.

Indeed, note that if $\varphi \in \mathcal{D}(\mathbb{K}^n)$, then naturally $\partial_i \varphi \in \mathcal{D}(\mathbb{K}^n)$ and therefore the right hand side of the definition is well defined. Furthermore, the map $\partial_i T : \varphi \longmapsto -\langle T, \partial_i \varphi \rangle$ is linear. In fact, for $\varphi, \psi \in \mathcal{D}(\mathbb{K}^n)$ and $\alpha \in \mathbb{K}$:

$$\begin{aligned} \langle \partial_i T, \varphi + \alpha \psi \rangle = -\langle T, \partial_i(\varphi + \alpha \psi) \rangle = -\langle T, \partial_i \varphi + \alpha \partial_i \psi \rangle = -(\langle T, \partial_i \varphi \rangle + \alpha \langle T, \partial_i \psi \rangle) \\ = -\langle T, \partial_i \varphi \rangle + \alpha(-\langle T, \partial_i \psi \rangle) = \langle \partial_i T, \varphi \rangle + \alpha \langle \partial_i T, \psi \rangle. \end{aligned}$$

The continuity of ∂_i follows from the fact that if $\varphi_j \xrightarrow{\mathcal{D}} 0$, then $\partial_i \varphi_j \xrightarrow{\mathcal{D}} 0$, and therefore $(-\langle T, \partial_i \varphi_j \rangle) \longrightarrow 0$, which shows that $\partial_i T \in \mathcal{D}'(\mathbb{K}^n)$.

One could ask if this is actually a generalization of the derivatives of usual functions. The answer however is already given in the motivation and moreover, the derivatives always commute. In fact, this result is given in the following proposition:

Proposition 7. *Let $f \in C^m(\mathbb{K}^n)$, then for every multi-index α such that $|\alpha| = m \in \mathbb{Z}_+$, $\partial^\alpha T_f = T_{\partial^\alpha f}$. Moreover, every m-order derivative of f with the same factors as α commute.*

Proof. The fact that $\partial^\alpha T_f = T_{\partial^\alpha f}$ has already been proven in the motivation for the definition of the derivatives of distributions. The fact that all distributions derivatives commute comes form the fact that, for every test function $\varphi \in \mathcal{D}(\mathbb{K}^n)$:

$$\langle \partial^\alpha T_f, \varphi \rangle = -\langle T_f, \partial^\alpha \varphi \rangle,$$

and, since $\varphi \in \mathcal{D}(\mathbb{K}^n) = C_c^\infty(\mathbb{K}^n)$ all derivatives of φ commute by Clairaut's Theorem. Hence, the derivatives of T_f also commute. $\square$

Example 15. *Let u be the Heaviside's step function on $\mathbb{R}$ and δ the Dirac's delta also "on $\mathbb{R}$" (because actually δ is not actually a real function). Then $u' = \delta$. In fact, this works in more generality, but as a first example, consider only the real case. Then, obviously $u \in L^1_{Loc}(\mathbb{R})$ and, for all test functions $\varphi \in \mathcal{D}(\mathbb{R})$, $\varphi(x) = 0$ for sufficiently large x. Therefore:*

$$u'(\varphi) = \langle u', \varphi \rangle = -\langle u, \varphi' \rangle = -\int_{-\infty}^{\infty} u \cdot \varphi' \, dx = -\int_0^\infty \varphi' \, dx = -\varphi(x)\Big|_0^\infty = \varphi(0) = \delta(\varphi).$$

Example 16. *The derivative δ' of the δ distribution is called the **dipole** and is given by:*

$$\langle \delta', \rangle = -\langle \delta, \varphi' \rangle = -\varphi'(0),$$

for all test functions $\varphi \in \mathcal{D}$.

Note that in Example 15 the classical derivative of u in the regions $x < 0$ and $x > 0$ is equal to 0 everywhere. The δ distribution, when thought as a "generalized function" as an arrow of infinite amplitude at $x = 0$ coincides with the classical derivative, except at $x = 0$, where the jump in the Heaviside function occurs. Philosophically, this infinite amplitude occurred exactly because of this jump.

Therefore, it is possible to write:

$$u' = \delta = T_0 + 1 \cdot \delta_0,$$

where T_0 is the distribution induced by the zero function and δ_0 corresponds to the "jump" in the value of u at $x = 0$. This is indeed no coincidence and is a particular case of the following result, that says roughly that the derivative of a function f in the sense of distributions is the classical derivative plus δ_a (delta distribution centered at $x = a$) times the jump of f at the point a where a jump occurs,

Proposition 8 (Jump Rule). *Let f be continuous differentiable on $\mathbb{R}$ except at the point $a \in \mathbb{R}$, where the limits $f(a+), f(a-), f'(a+), f'(a-)$ exist. Then f, f' are locally integrable and*

$$(T_f)' = T_{f'} + (f(a+) - f(a-))\delta_a,$$

where delta$_a$ is the delta distribution centered at $x = a$.

Proof. Let $\varphi \in \mathcal{D}(\mathbb{R})$ and suppose that φ is 0 outside a interval $[\alpha, \beta]$ such that $a \in [\alpha, \beta]$. Then:

$$\begin{aligned}
\langle (T_f)', \varphi \rangle &= -\langle T_f, \varphi' \rangle \\
&= -\int_\alpha^\beta f(x)\varphi'(x)dx \\
&= -\int_\alpha^a f(x)\varphi'(x)dx - \int_a^\beta f(x)\varphi'(x)dx \\
&= \int_\alpha^a f'(x)\varphi(x)dx - f(a-)\varphi(a) + \int_a^\beta f'(x)\varphi(x)dx + f(a+)\varphi(a) \\
&= \int_\alpha^\beta f'(x)\varphi(x)dx + (f(a+) - f(a-))\varphi(a) \\
&= \langle T_{f'}, \varphi \rangle + (f(a+) - f(a-))\langle \delta_a, \varphi \rangle \\
&= \langle T_{f'} + (f(a+) - f(a-))\delta_a, \varphi \rangle.
\end{aligned}$$

□

Observation 3. *The previous result can be immediately extended to the case where f is continuously differentiable except on a finite number of points a_k such that at these points the function satisfies the same assumptions required before. In this case:*

$$(T_f)' = T_{f'} + \sum_k (f(a_k+) - f(a_k-))\delta_{a_k}.$$

Furthermore, this result can actually be extended to the case where f has infinitely many, but locally finite, jump discontinuities, i.e., in any compact interval one can only find finitely many of these discontinuities. In this case, denoting $\sigma_k = (f(a_k+) - f(a_k-))$, the sum on the right hand side will be the distribution defined by:

$$\left\langle \sum_k \sigma_k \delta_{a_k}, \varphi \right\rangle := \sum_k \sigma_k \varphi(a_k), \quad \varphi \in \mathcal{D}(\mathbb{R})$$

where, for a given test function φ, only finitely many terms on the right hand side are nonzero.

To finish this section, it is important to recall from the beginning that one motivation in the development of the theory of distributions was indeed to generalize the notion of derivatives and for this to be useful in many mathematical areas. Sometimes this derivatives in the sense of distributions are called **weak derivatives**.

To illustrate this fact, one can consider examples studied in Partial Differential Equations (P.D.E).

Definition 25. *A distribution satisfying a P.E.D. in the sense of distributions will be called a* ***weak solution*** *to a P.E.D.*

Finally, this is illustrated in the following example that is very important in physics and engineering:

Example 17 (Weak Solution To The Wave Equation). *Recall that given a function $f \in C^2(\mathbb{R})$, then*

$$u(x,t) := \frac{f(x+t)+f(x-t)}{2},$$

is a classical solution to the ***wave equation***

$$\frac{\partial^2 u}{\partial t^2} - \frac{\partial^2 u}{\partial x^2} = 0,$$

given boundary conditions $u(x,0) = f(x)$ and with zero initial speed, i.e., $u_t(x,0)$.

Indeed, if the function f is only locally integrable, the u expressed above is still a solution to the P.E.D., however, now in the sense of distributions.

To prove this, the following result concerning the ***transport equation*** *is useful (see [18] for this). Given $f \in L^1_{Loc}(\mathbb{R})$, then u_+ and u_- given by*

$$\begin{cases} u_+(x,t) := f(x+t), \\ u_-(x,t) := f(x-t), \end{cases}$$

are weak solutions to

$$\begin{cases} \dfrac{\partial u}{\partial t} - \dfrac{\partial u}{\partial x} = 0, \\ \dfrac{\partial u}{\partial t} + \dfrac{\partial u}{\partial x} = 0, \end{cases}$$

respectively.

Now recall that for a distribution $T \in \mathcal{D}'(\mathbb{R}^2)$:

$$\frac{\partial^2 T}{\partial x \partial t} = \frac{\partial^2 T}{\partial t \partial x}.$$

Hence:

$$\left(\frac{\partial}{\partial t} - \frac{\partial}{\partial x}\right)\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial x}\right) T = \left(\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x^2}\right) T = \left(\frac{\partial}{\partial t} + \frac{\partial}{\partial x}\right)\left(\frac{\partial}{\partial t} - \frac{\partial}{\partial x}\right) T.$$

So let u be given by

$$u(x,t) = \frac{f(x+t) + f(x-t)}{2},$$

and $\varphi \in \mathcal{D}(\mathbb{R}^2)$. Then:

$$\begin{aligned} \left\langle \frac{\partial^2 u}{\partial t^2} - \frac{\partial^2 u}{\partial x^2}, \varphi \right\rangle &= \frac{1}{2}\left\langle \left(\frac{\partial^2 u}{\partial t^2} - \frac{\partial^2 u}{\partial x^2}\right) f(x+t), \varphi \right\rangle + \frac{1}{2}\left\langle \left(\frac{\partial^2 u}{\partial t^2} - \frac{\partial^2 u}{\partial x^2}\right) f(x-t), \varphi \right\rangle \\ &= \frac{1}{2}\left\langle \left(\frac{\partial}{\partial t} + \frac{\partial}{\partial x}\right)\left(\frac{\partial}{\partial t} - \frac{\partial}{\partial x}\right) f(x+t), \varphi \right\rangle + \frac{1}{2}\left\langle \left(\frac{\partial}{\partial t} - \frac{\partial}{\partial x}\right)\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial x}\right) f(x-t), \varphi \right\rangle \\ &= -\frac{1}{2}\left\langle \left(\frac{\partial}{\partial t} - \frac{\partial}{\partial x}\right) f(x+t), \left(\frac{\partial}{\partial t} + \frac{\partial}{\partial x}\right)\varphi \right\rangle - \frac{1}{2}\left\langle \left(\frac{\partial}{\partial t} + \frac{\partial}{\partial x}\right) f(x-t), \left(\frac{\partial}{\partial t} - \frac{\partial}{\partial x}\right)\varphi \right\rangle \\ &\overset{\star}{=} -0 - 0 = 0, \end{aligned}$$

where the $(\star)$ must be justified. To see that this is indeed the case, notice that for $\varphi \in \mathcal{D}(\mathbb{R}^2)$, then

$$\left(\frac{\partial}{\partial t} \pm \frac{\partial}{\partial x}\right)\varphi \in \mathcal{D}(\mathbb{R}^2).$$

But for any test function ψ

$$\left\langle \left(\frac{\partial}{\partial t} - \frac{\partial}{\partial x}\right) f(x+t), \psi \right\rangle = 0 = \left\langle \left(\frac{\partial}{\partial t} + \frac{\partial}{\partial x}\right) f(x-t), \psi \right\rangle,$$

and, therefore, $(\star)$ holds.

2.10 Mollifiers And The Density Of Regular Distributions

In this section, the goal will be to show how any distribution can be approximated by regular distributions and, therefore, in particular justifying the definition for the duality pairing.

However, firstly, it is important to define the *Mollifiers* and the process of *mollification* to show how a distribution can by approximated by a sequence of smooth functions.

Definition 26. *Let $U \subseteq M_n$ be an open subset and $\varphi \in \mathcal{D}(U)$ satisfying the following properties:*

1. $\displaystyle\int_{M_n} \varphi \, d\lambda = 1$*;*
2. $\displaystyle\lim_{\varepsilon \longrightarrow 0} \varphi_\varepsilon(x) = \lim_{\varepsilon \longrightarrow 0} \varepsilon^{-n} \varphi\left(\frac{x}{\varepsilon}\right) = \delta(x),$

*where the δ represents the Dirac's Delta and the limit must be understood in the space of distributions. Then the function φ is called a **Mollifier** or an **Approximation Of The Identity**.*

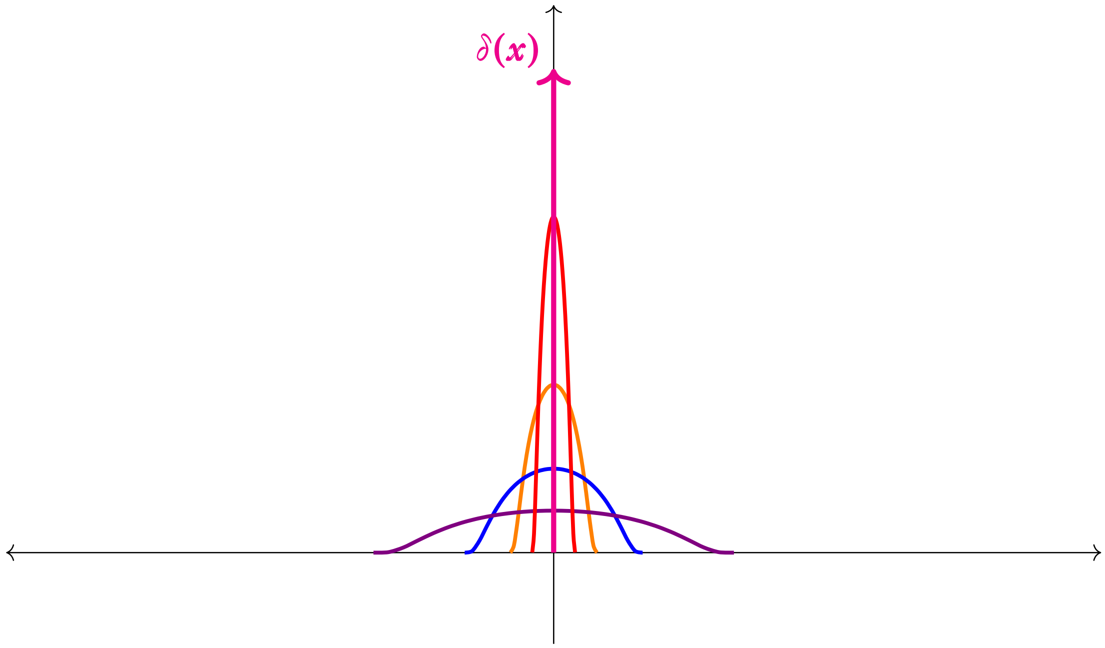

Figure 2.9: Mollifiers approximating the Dirac's Delta.

Consider here $M_n = \mathbb{R}^n$ and consider a function $\eta \in C(\mathbb{R}^n)$ a non-negative function such that $\operatorname{supp}(\eta) \subseteq B_1 = \{x \in \mathbb{R}^n : \|x\| < 1\}$. Moreover let η satisfy that $\int_{\mathbb{R}^n} \eta \, d\lambda = 1$.

For example, taking $\eta = \psi$ the bump function:

$$\eta(x) = \begin{cases} c \exp\left(\dfrac{-1}{1 - \|x\|^2}\right), & \text{for } \|x\|^2 < 1, \\ 0, & \text{otherwise.} \end{cases}$$

The constant c is just to ensure that the integral is unitary.

Definition 27. *Let $\Omega \subseteq \mathbb{R}^n$ be a bounded domain (i.e., non-empty connected open set) and $f \in L^1(\Omega)$. Given $\varepsilon > 0$, the* ***mollification*** *of f, denoted by f_ε, is defined by:*

$$f_\varepsilon(x) = \varepsilon^{-n} \int_\Omega \eta\left(\frac{x-y}{\varepsilon}\right) f(y) \, dy \quad \text{for } x \in \Omega \text{ and } \varepsilon < dist(x, \partial\Omega);$$

where $\partial\Omega$ is the boundary of Ω.

Notation 2. *Let the symbol $\Subset$ be used in $A \Subset B$ to denote that $\overline{A} \subseteq B$.*

Lemma 2. *If $\tilde{\Omega} \Subset \Omega$ and $0 < \varepsilon < dist(\tilde{\Omega}, \partial\Omega)$, then $f_\varepsilon \in C^\infty(\tilde{\Omega})$.*

Proof. See [12]. □

Just as in the notation at the beginning of this chapter, denote for $k \in \mathbb{Z}_+ \cup \{\infty\}$:

$$C_c^k(\Omega) = \{\varphi \in C^k(\Omega) : \operatorname{supp}(\varphi) \Subset \Omega\}.$$

The following result ensures one of the central aspects of this subject.

Theorem 7. *Let $f \in C^0(\Omega)$. Then, the mollification f_ε converges uniformly on compact sets to f as ε goes to zero, i.e.,*

$$f \xrightarrow{\varepsilon \longrightarrow 0} f_\varepsilon.$$

Proof. By definition of f_ε:

$$f_\varepsilon(x) = \varepsilon^{-n} \int\limits_{\|x-x_0\|<\varepsilon} \eta\left(\frac{x-y}{\varepsilon}\right) f(y)\, dy = \int\limits_{\|z\|<1} \eta(z) f(x-\varepsilon z)\, dz,$$

where $z = (x-y)/\varepsilon$ and x_0 its just a point that translate the ε-ball to the right place. Now, if $\tilde{\Omega}$ is such that $\tilde{\Omega} \Subset \Omega$ and $\varepsilon < (1/2) \cdot \operatorname{dist}(\tilde{\Omega}, \partial\Omega)$, since

$$f(x) = \int\limits_{\|z\|<1} \eta(z) f(x)\, dz,$$

then

$$
\begin{aligned}
\sup_{x\in\tilde{\Omega}} |f(x) - f_\varepsilon(x)| &= \sup_{x\in\tilde{\Omega}} \left| \int_{\|z\|\le 1} \eta(z)\cdot(f(x) - f(x-\varepsilon z))\, dz \right| \\
&\le \sup_{x\in\tilde{\Omega}} \int_{\|z\|\le 1} \left|\eta(z)\cdot(f(x) - f(x-\varepsilon z))\right|\, dz \\
&= \sup_{x\in\tilde{\Omega}} \int_{\|z\|\le 1} \eta(z) \left|\cdot(f(x) - f(x-\varepsilon z))\right|\, dz \\
&\le \sup_{x\in\tilde{\Omega}} \sup_{\|z\|\le 1} \left|(f(x) - f(x-\varepsilon z))\right|.
\end{aligned}
$$

Since f is continuous in the compact set Ω, it will also be uniformly continuous, then

$$
\lim_{\varepsilon\longrightarrow 0}\left(\sup_{x\in\tilde{\Omega}} \sup_{\|z\|\le 1} \left|(f(x) - f(x-\varepsilon z))\right|\right) = 0.
$$

Therefore, because f is uniformly continuous on Ω' (the compact set $\overline{\operatorname{supp}(f)} \subseteq \Omega \supseteq \Omega'$ and $f \equiv 0$ outside Ω) the upperbound above tends to zero as ε goes to zero. □

For the next result, the following theorem is useful.

Theorem 8 (Lusin's). *Let μ be a Borel regular measure on $\mathbb{R}^n$ and $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ be a μ-measurable function. Also, let $A \subseteq \mathbb{R}^n$ be a μ-measurable set with $\mu(A) < \infty$. Then, for a fixed $\varepsilon > 0$, there exists a compact set $K \subseteq A$ such that*

1. $\mu(A \setminus K) < \varepsilon$,

2. $f\big|_K$ *is continuous (and therefore uniformly continuous).*

Proof. See [7]. □

As a straightforward corollary:

Corollary 4. *Let* $\Omega \subseteq \mathbb{R}^n$ *be a bounded domain and* $f \in L^p(\Omega)$, $1 \leq p < \infty$ *and* $\varepsilon > 0$. *Then, there exists* $g \in C^0(\mathbb{R}^n)$ *such that* $\|f - g\|_p < \varepsilon$.

Proof. This is an obvious consequence of Theorem 8. □

Theorem 9. *Let* $\Omega \subseteq \mathbb{R}^n$ *be a bounded domain and* $f \in L^p(\Omega)$, $1 \leq p < \infty$. *Then, extending* f *as the zero function in* $\mathbb{R}^n \setminus \Omega$, f_ε *can be defined in all* Ω *and*

$$f_\varepsilon \xrightarrow[\varepsilon \longrightarrow 0]{L^p} 0.$$

Proof. Starting just as in Theorem 7,

$$\begin{aligned}
f_\varepsilon(x) &= \int\limits_{\|z\| \leq 1} \eta(z) f(x - \varepsilon z)\, dz \\
&= \int\limits_{\|z\| \leq 1} \left(\eta(z)^{1-\frac{1}{p}}\right) \cdot \left(\eta(z)^{1/p} f(x - \varepsilon z)\right)\, dz \\
&\leq \left(\int\limits_{\|z\| \leq 1} \eta(z)\, dz \right) \cdot \left(\int\limits_{\|z\| \leq 1} \eta(z) \cdot |f(x - \varepsilon z)|^p \right)^{\frac{1}{p}} \\
&\overset{\star}{=} \left(\int\limits_{\|z\| \leq 1} \eta(z) \cdot |f(x - \varepsilon z)|^p\, dz \right)^{\frac{1}{p}},
\end{aligned}$$

where ($\star$) comes from the definition of η. Hence:

$$
\begin{aligned}
\int\limits_{\Omega} |f_\varepsilon(x)|^p \, dx &\leq \int\limits_{\Omega} \int\limits_{\|z\|\leq 1} \eta(z)|f(x-\varepsilon z)|^p \, dz \, dx \\
&\overset{\star\star}{\leq} \int\limits_{\|z\|} \eta(z) \int\limits_{\Omega} \left(|f(x-\varepsilon z)|^p \, dx\right) \, dz \\
&\leq \int\limits_{\|z\|\leq 1} \eta(z) \left(\int\limits_{\mathbb{R}^n} |f(y) \, dy| \right) dz \\
&= \int\limits_{\mathbb{R}^n} |f(y)|^p \, dy
\end{aligned}
\tag{2.2}
$$

where ($\star\star$) comes from Fubini's Theorem.

Therefore, for a given $\varepsilon > 0$. Thus, take $\tilde{\Omega} \Subset \Omega$ such that:

$$
\left(\int\limits_{\Omega \setminus \tilde{\Omega}} |f(x) - f_\varepsilon(x)| \, dx \right)^{\frac{1}{p}} < \frac{\varepsilon}{4}, \tag{2.3}
$$

which can be done due to Corollary 4. Hence, if $\psi \in L^p(\Omega)$, then for every $\delta > 0$, there will exist $\Omega' \Subset \Omega$ such that

$$
\int\limits_{\Omega \setminus \Omega'} |\psi(x)|^p \, dx < \delta. \tag{2.4}
$$

Then, take $\varphi \in C^0(\mathbb{K}^n)$ such that

$$
\|f - \varphi\|_p < \frac{\varepsilon}{4}. \tag{2.5}
$$

By Theorem 7, and reinforcing that the Lebesgue measure λ is being used, for sufficiently small ε:

$$
\|\varphi - \varphi_\varepsilon\|_{L^p(\mathbb{K}^n)} \leq \lambda(\Omega)^{\frac{1}{p}} \cdot \sup_{x \in \tilde{\Omega}} |\varphi(x) - \varphi_\varepsilon(x)| \leq \frac{\varepsilon}{4}. \tag{2.6}
$$

Repeating (2.2) but with $(f-\varphi)$ instead of f:

$$\int_{\Omega} |\varphi_{\varepsilon}(x) - f_{\varepsilon}(x)|^p \, dx \leq \int_{\Omega} |\varphi(x) - f(x)|^p \, dx. \tag{2.7}$$

Then:

$$\begin{aligned}
\|f - f_{\varepsilon}\|_p &\leq \|f - f_{\varepsilon}\|_{L^p(\tilde{\Omega})} + \|f - f_{\varepsilon}\|_{L^p(\Omega \setminus \tilde{\Omega})} \\
&\leq \|f - \varphi\|_{L^p(\mathbb{R}^p)} + \|\varphi - \varphi_{\varepsilon}\|_{L^p(\tilde{\Omega})} + \|\varphi_{\varepsilon} - f_{\varepsilon}\|_{L^p(\mathbb{R}^n)} + \frac{\varepsilon}{4} \\
&< \varepsilon.
\end{aligned}$$

□

With these results, one can already establish the results about the density of $\mathcal{D}(\Omega)$ in $L^p(\Omega)$.

Theorem 10. *$\mathcal{D}(\Omega)$ is dense in $L^p(\Omega)$ with $1 \leq p < \infty$, i.e., for every $f \in L^p(\Omega)$ and a given $\varepsilon > 0$, there exists $\varphi \in \mathcal{D}(\Omega)$ with $\|f - \varphi\|_{L^p(\Omega)} < \varepsilon$.*

Proof. Just as before, star choosing an appropriate $\tilde{\Omega} \Subset \Omega$ such that:

$$\|\|_{L^p(\Omega \setminus \tilde{\Omega})} \leq \frac{\varepsilon}{3},$$

and define:

$$g(x) = \begin{cases} f(x), & \text{if } x \in \overline{\tilde{\Omega}}, \\ 0, & \text{if } x \in \mathbb{K}^n \setminus \overline{\tilde{\Omega}} \end{cases}$$

Hence by Theorem 9, there exists $\varepsilon < \operatorname{dist}(\Omega, \tilde{\partial}\Omega)$ such that:

$$\|g - g_{\varepsilon}\|_{L^p(\Omega)} < \frac{\varepsilon}{3}.$$

Now, since $g \equiv 0$ in $\Omega \setminus \tilde{\Omega}$, then

$$\|g_\varepsilon\|_{L^p(\Omega\setminus\tilde{\Omega})} < \frac{\varepsilon}{3}.$$

Therefore:

$$\|f - g_\varepsilon\|_{L^p(\Omega)} \leq \|f\|_{L^p(\Omega\setminus\tilde{\Omega})} + \|g\|_{L^p(\Omega\setminus\tilde{\Omega})} + \|g - g_\varepsilon\|_{L^p(\Omega)} < \varepsilon.$$

Hence, by Lemma 2, $g_\varepsilon \in C^\infty(\mathbb{K}^n)$, and, by the choose of the ε above, g_ε has compact support in Ω.

□

Now that the fact that $\mathcal{D}(\Omega)$ is dense in $L^p(\Omega)$, one could ask what happens in the space $\mathcal{D}'(\Omega)$, *i.e.*, are the regular distributions dense in this space? In fact, this is also true, meaning that test functions can be viewed as a dense subset of both $L^p(\Omega)$ and $\mathcal{D}'(\Omega)$. This means in particular that every distribution can be approximated by regular ones, justifying many of the definitions made above such as the derivatives, coordinate changes, etc.

The proof of this second fact concerning the density of test functions (actually the distributions induced by these test functions) in the space of distributions, however, requires some more tools. In particular the definitions of the *support of a distribution* and the *convolution of distributions* will be used.

Definition 28. *Let $T \in \mathcal{D}'(\Omega)$, then the* ***support of the distribution*** *T is defined as supp$(T) = \{x \in \Omega :$ There is no neighbourhood of x with $T \equiv 0$ in this neighbourhood$\}$.*

The Definition 28 above may seem a bit strange since distributions are not functions. More precisely, if $f : \mathbb{R}^n \longrightarrow \mathbb{C}$ is a function, the support of f is $\mathrm{supp}(f) = \overline{\{x \in \mathbb{R}^n \ : \ f(x) \neq 0\}}$ and moreover, it does make sense to talk about the value of f at a certain point x. In the case of a distribution, it does not make sense to evaluate T at some point x. Nonetheless, since distributions are linear functionals, the distribution could be interpreted as having value zero at a set Λ where it does not do anything, *i.e.*, where $\langle T, \varphi \rangle = 0$ for every φ such that $\mathrm{supp}(\varphi) \subseteq \Lambda$.

It is not very hard to prove that the support of a distribution is a closed set. The following result is also interesting.

Proposition 9. *Let $\varphi \in \mathcal{D}(\Omega)$ and $T \in \mathcal{D}'(\Omega)$. If supp$(\varphi) \cap$ supp$(T) = \emptyset$, then $T(\varphi) = 0$.*

Proof. See [5]. $\square$

In fact, to prove the Proposition 9, [5] uses some interesting results about the local behaviour of distributions which are very interesting and proven using partitions of the unity.

Now, if $T = T_f$ for some $f \in L^1_{\text{Loc}}(\mathbb{R}^n)$ and $\varphi \in \mathcal{D}(\mathbb{R}^n)$, then

$$[f * \varphi] = \int_{\mathbb{R}^n} f(y)\varphi(x-y)\,dy.$$

This fact motivates the following definition:

Definition 29. *Let $T \in \mathcal{D}(\mathbb{R}^n)$ and $\varphi \in \mathcal{D}(\mathbb{R}^n)$. Then the **convolution** of a distribution with a test function is defined as*

$$[T * \varphi](x) := \langle T, \varphi_x \rangle,$$

where $\varphi_x : y \longmapsto \varphi(x-y)$.

Although distributions are not functions, the convolution of a distribution with a test function is often written as:

$$[T * \varphi](x) := \langle T(y), \varphi(x-y) \rangle,$$

in the meaning that the distribution T acts on the test function $y \longmapsto \varphi(x-y)$. Sometimes this is also written as $\langle T, \varphi(x-\cdot)\rangle$ or $\langle T_y, \varphi(x-y)\rangle$.

Theorem 11. *If $T \in \mathcal{D}'(\mathbb{R}^n)$ and $\varphi \in \mathcal{D}(\mathbb{R}^n)$, then*

1. *$T * \varphi \in C^\infty(\mathbb{R}^n)$;*

2. supp(T ∗) ⊆ supp(T) ∪ supp();

3. $\partial^\alpha(T * \varphi) = T * \partial^\alpha \varphi = \partial^\alpha(T) * \varphi$.

Proof. Firstly, it is important to show that $T * \varphi$ is continuous. Take points x such that $x \longrightarrow x_0$. If $|x - x_0| \leq 1$, then $y \longmapsto \varphi(x - y)$ has its support in a fixed compact set. Moreover:

$$\partial_y^\alpha(\varphi(x-y) - \varphi(x_0 - y)) \xrightarrow[\text{unif}]{x \longrightarrow x_0} 0,$$

where the notation ∂_y^α simply means that the multi-index derivative is taken with respect to the variable $y = (y_1, \ldots, y_n)$. Therefore, $\varphi(x - y) \xrightarrow{\mathcal{D}} \varphi(x_0 - y)$ as $x \longrightarrow x_0$ and, hence:

$$[T * \varphi](x) = \langle T_y \varphi(x - y) \rangle \xrightarrow{x \longrightarrow x_0} \langle T_y, \varphi(x_0 - y) \rangle = [T * \varphi](x_0).$$

Now that the fact that $[T * \varphi]$ is continuous is proven, in order to prove (2), it suffices to show that if $x \notin \operatorname{supp}(T) \cup \operatorname{supp}(\varphi)$, then $T * \varphi(x) = 0$. Suppose then that $x \notin \operatorname{supp}(\varphi) \cup \operatorname{supp}(T)$. Then, there are no $y \in \operatorname{supp}(T)$ such that $(x - y) \in \operatorname{supp}(\varphi)$ and therefore there is no y with $y \in \operatorname{supp}(T)$ and $y \in \operatorname{supp}(\varphi(x - \cdot))$. Hence, $\operatorname{supp}(T) \cap \operatorname{supp}(\varphi) = \emptyset$, which implies that $T * \varphi(x) = 0$.

The second equality in (3) can be proved observing that:

$$\partial^\alpha T * \varphi(x) = \langle \partial^\alpha T_y, \varphi(x - y) \rangle = (-1)^{|\alpha|} \langle T_y, \partial_y^\alpha \varphi(x - y) \rangle = T * (\partial^\alpha \varphi)(x).$$

The first equality in (3) however can be proved by induction taking as base case $\alpha = (1, 0, \ldots, 0) = e_1$. Hence, it suffices to show that:

$$\lim_{h \longrightarrow 0} \frac{1}{h}(T * \varphi(x + he_1) - T * \varphi(x)) = T * \partial_1 \varphi(x).$$

Thus, let

$$\phi_{x,h}(y) = \frac{1}{h} \cdot \left(\varphi(x + he_1 - y) - \varphi(x - y)\right).$$

Then

$$\frac{1}{h} \cdot \left(T * \varphi(x + he_1) - T * \varphi(x)\right) = T\left(\phi_{x,h}\right)$$

However $\phi_{x,h}(y) \xrightarrow{\mathcal{D}} \frac{\partial \varphi}{\partial x_1}$. Therefore:

$$\partial^\alpha [T * \varphi](x) = \lim_{h \longrightarrow 0} T(\phi_{x,h}) = T_y\left(\frac{\partial \varphi}{\partial x_1}(x - y)\right) = T * \partial_1 \varphi(x).$$

Since (1) follows from (3), the theorem is proved. □

Proposition 10. *Let $T \in \mathcal{D}'(\mathbb{R}^n)$ and $\varphi, \psi \in \mathcal{D}(\mathbb{R}^n)$. Then $(T * \varphi) * \psi = T * (\varphi * \psi)$.*

Proof. Indeed:

$$\begin{aligned} T * (\varphi * \psi)(x) &= \langle T_y, \varphi * \psi(x - y)\rangle = \left\langle T_y, \int_{\mathbb{R}^n} \varphi(x - y - t)\psi(t)\, dt \right\rangle \\ &\overset{\star}{=} \int_{\mathbb{R}^n} \langle T_y, \varphi(x - y - t)\rangle \psi(t)\, dt = \int_{\mathbb{R}^n} T * \varphi(x - t)\psi(t)\, dt \\ &= (T * \varphi) * \psi(x) \end{aligned}$$

where the equality ($\star$) must be justified. In fact that result holds and can be proved using approximations to the Riemann Integral. This is a consequence of the following lemma, which ensures that the Riemann sums converges in $\mathcal{D}$ to the convolution and, hence, ($\star$) is valid. □

Lemma 3. *If $\varphi \in C_c^j(\mathbb{R}^n)$ and $\varphi \in C_c(\mathbb{R}^n)$, then*

$$\sum_{k \in Z^n} \varphi(x - kh)\psi(kh)h^n \xrightarrow{h \longrightarrow 0} \varphi * \psi(x),$$

in C_c^j (and the convergence in C_c^j is defined in the same fashion as the $\mathcal{D}$ convergence mutatis mutandis).

Proof. This sum is supported in $\operatorname{supp}(\varphi) \cup \operatorname{supp}(\psi)$. Moreover, the map $(x, y) \longmapsto \varphi(x-y)\psi(y)$ is uniformly continuous. Therefore, the Riemann sums converge uniformly to $\varphi * \psi(x)$. And since $\partial^\alpha(\varphi * \psi) = \partial^\alpha \varphi * \psi$ for $|\alpha| \leq j$, this is also true for the derivatives. □

With these results, the final conclusion about the density of regular distributions in the space of distributions can be proved:

Theorem 12. *Let $T \in \mathcal{D}(\mathbb{R}^n)$ and φ_ε be a mollifier. Then:*

$$T * \varphi_\varepsilon \xrightarrow[\varepsilon \longrightarrow 0]{\mathcal{D}'} T.$$

Proof. Define the function $\check{\psi}$ as $\check{\psi}(x) = \psi(-x)$. Then $T(\psi) = T * \check{\psi}(0)$. Hence, Proposition 10 implies that

$$T_\varepsilon(\psi) = T * \varphi_\varepsilon(\psi) = (T * \varphi_\varepsilon) * \check{\psi}(0) = T * (\varphi_\varepsilon * \check{\psi})(0).$$

However, φ_ε is a mollifier, and therefore

$$\varphi_\varepsilon * \check{\psi} \xrightarrow[\varepsilon \longrightarrow 0]{\mathcal{D}} \check{\psi}.$$

Hence:

$$\lim_{\varepsilon \longrightarrow 0} T_\varepsilon(\psi) = \lim_{\varepsilon \longrightarrow 0} T * (\varphi_\varepsilon * \check{\psi})(0) = T * \check{\psi}(0) = T(\psi).$$

□

Corollary 5. *If $T \in \mathcal{D}'(\Omega)$, there are functions T_n in $\mathcal{D}$ such that*

$$T_n \xrightarrow[n \longrightarrow \infty]{\mathcal{D}'} T.$$

Proof. Immediate from the previous result given. It suffices to take $\varepsilon = 1/n$ at each step. $\square$

This shows once again that the spaces $\mathcal{D}$ and $\mathcal{D}'$ are indeed very good in regularity to work on. In particular, $\mathcal{D}$ is dense both in L^p and $\mathcal{D}'$ (of course considering the isomorphism given by the functions and the induced distributions).

It is by logic that we prove, but by intuition that we discover.

To know how to criticize is good, to know to create is better.

Henri Poincaré

3

Schwartz Space, Tempered Distributions And Fourier Transform

The usual Fourier Transform is defined in the space of square-integrable functions, *i.e.*, is a transformation of the form $\mathcal{F} \ : \ L^2(\mathbb{R}) \longmapsto L^2(\mathbb{R})$. One might want to do the same with distributions in the space $\mathcal{D}'$ in hopes of extending this concept. However, this does not work immediately.

In a first attempt to define the Fourier Transform of a distribution $T \in \mathcal{D}'$, $\varphi \longmapsto T(\varphi) \in \mathbb{K}$, one might assume $\mathbb{K} = \mathbb{R}$ and try to use the same formula and, instead of a function with frequency $\omega \in \mathbb{R}$, this could lead to a "frequency" that is a test function:

$$\hat{T}(\psi) = \mathcal{F}[T](\psi) = \int_{-\infty}^{\infty} e^{i\psi\varphi} T(\varphi)\, d\varphi,$$

but this is not at all well defined staring with the fact that $e^{i\psi\varphi}$ is not a test function in $\mathcal{D}$, so the action of T on $e^{i\psi\varphi}$ is not well defined.

For a second attempt, one could try Parseval's formula from classical real and complex analysis:

$$\int_{-\infty}^{\infty} \hat{f}(x) g(x)\, dx = \int_{-\infty}^{\infty} f(x)\hat{g}(x)\, dx,$$

which is a formula that connects the Fourier Transform of two functions f and g. Thus, one natural attempt would be to try to define:

$$\langle \hat{T}, \varphi \rangle = \langle T, \hat{\psi} \rangle, \qquad \varphi \in \mathcal{D}.$$

However, this does not work as well since the Fourier Transform of a test function may not be a test function.

The way this is actually done is to enlarge the set of test functions by only requiring that they have "decay faster then every polynomial at infinity", which will be made more precise later (they will also be smooth, but this is just as was before). The distributions will also be different, since they will be continuous linear functionals defined on this new set. This new type of "test functions" will be called *test functions of rapid decay* and the set of those functions the *Schwartz Space* ($\mathcal{S}$), whilst these new distributions will be named *tempered distributions* and form a set $\mathcal{S}$. Since it is obvious that the compact smooth functions decay faster as one approaches infinity then every polynomial (since these functions are identically zero outside a compact set), it is immediate that $\mathcal{D} \subseteq \mathcal{S}$, and hence $S' \subseteq D'$.

This may seem confusing but it is logical since a functional that is linear and continuous in a bigger set will necessarily be linear and continuous in a smaller subset. This restriction is done because as said before, Fourier Transforms are not defined for every distribution, but, in the space $\mathcal{S}'$ it will be a linear automorphism, which makes the Schwartz Space a very natural space to work with Fourier Transforms.

3.1 Schwartz Space And Tempered Distributions:

Before getting into the new concepts, a few notations must be recapitulated.

Denote by $\Omega \subset \mathbb{R}^n$ an open subset. A multi-index $\alpha = (\alpha_1, \ldots, \alpha_n)$ with $\alpha_i \in \mathbb{Z}_+$ will be used for all $1 \leq i \leq n$. The order of a multi-index α is $|\alpha| = \sum_{i=1}^{n} \alpha_i$.

A few more notation conventions will be used to make the reading more fluid:

- $\alpha! = \alpha_1 \cdot \alpha_2 \cdots \alpha_n$
- $x^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n}$
- $\partial^\alpha = \partial_{x_1}^{\alpha_1} \cdots \partial_{x_n}^{\alpha_n}$

The first step is to define this new space of test functions more precisely:

Definition 30. *The Schwartz's Space, $S(\mathbb{R}^n)$, is the set of functions $f \in C^\infty(\mathbb{R}^n)$ such that:*

$$||f||_{\alpha\beta} = sup|x^\alpha \partial^\beta f(x)| < C_{\alpha\beta} < \infty,$$

is valid for every α and β in $(\mathbb{Z}_+)^n$, where $C_{\alpha\beta}$ is a constant depending on α, β and f.

The above definition may be interpreted as the set of smooth functions of rapid decay, *i.e.*, they vanish at infinity faster than the reciprocal of any polynomial.

Example 18. *As spoiled before at the beginning of this chapter, the space* $\mathcal{D}(\mathbb{R}^n) = C_c^\infty(\mathbb{R})$ *of the usual test functions (smooth and with compact support) is a subspace of* $\mathcal{S}(\mathbb{R})^n$*. Since those functions vanish outside a compact (which will be closed and bounded since* $\mathbb{R}^n$ *here is a metric space in the usual topology),* $|x^\alpha \partial^\beta f(x)|$ *will assume a finite maximum in* $\mathbb{R}^n$ *(Stone-Weierstrass's Theorem).*

The previous example showed that $\mathcal{D} \subseteq \mathcal{S}$. However, this inclusion is strict, *i.e.*, $\mathcal{D} \subsetneq \mathcal{S}$, as shown in the next example:

Example 19. *Take* $n = 1$ *and consider the real function* $f = e^{-x^2}$*. Note that* $|x^\alpha f^{(n)}(x)|$ *is always of the form* $|p(x)e^{-x^2}|$ *with* $p(x)$ *being a polynomial. Taking the limit as* $|x| \to \infty$*, this results in* 0 *and, hence,* f *belongs to the Schwartz's Space. However it is obvious that* f *is no compactly supported, because* $f \neq 0$ *for all* $x \in \mathbb{R}$*.*

One of the most interesting things about Example 18 is that it implies that $\mathcal{S}(\mathbb{R}^n)$ is dense in $L^p(\mathbb{R}^n), 1 \leq p < \infty$. If one remembers from the last chapter that indeed $\mathcal{D}$ is dense in $L^p, 1 \leq p < \infty$, this is straightforward. However, this can be proved without that fact. To see why this is the cases, it suffices to remember that $C^\infty(\mathbb{R}^n)$ is dense in $L^p(\mathbb{R}^n)$. However, first of all, it needs to be proven that this claim in fact makes sense, *i.e.*, that $\mathcal{S}(\mathbb{R}^n) \subset L^p(\mathbb{R}^n)$.

Observation 4. *Before proceeding with the reasoning, it is important to make a remark about another characterization of the space* $\mathcal{S}(\mathbb{R}^n)$*. Firstly, note that if* f *is dominated by a polynomial* $x^m = x_1^{m_1} x_2^{m_2} \cdots x_n^{m_n}$*, then, choosing* $N = \max\{m_1, \ldots, m_n\}$*, writing in a abuse of notation* $x^N = x_1^N \cdots x_n^N$*, then* f *is also dominated by* x^N*. Moreover, if* N *is even, then changing* x *for* $|x|$ *will not change anything, while if* N *is odd,* $N+1$ *is even and* x^{N+1} *dominates* x^N*, and hence dominates* f*. Hence, another characterization of the Schwartz Space is that* $|\partial^\alpha f(x)| \leq C_{N,\alpha} \cdot (|x|)^N$ *for every positive integer* N*. Moreover, sometimes is easier for purposes of calculation to state that* $|\partial^\alpha f(x)| \leq C_{N,\alpha} \cdot (1 + |x|)^N$*, which does not change the result.*

Lemma 4. *Let* $1 \leq p < \infty$. *Then,* $S(\mathbb{R}^n) \subset L^p(\mathbb{R}^n)$.

Proof. As remarked before, another characterization of $S(\mathbb{R}^n)$ is that: $|\partial^\alpha f(x)| \leq C_{N,\alpha} \cdot (1+|x|)^{-N}$, for N positive integer (also in the abuse of notation, since is a multi-index with n copies of N). It is possible to take α as the zero vector, *i.e.*, take the module of the function, which makes sense since it is already known that the derivatives go to zero as $|x|$ grows. Hence, if f was not bounded, it would have to be unbounded in a region that is not the extremes and thus, its derivatives could not be bounded, a contradiction. Hence it needs to be verified that $f \in L^p$ for $1 \leq p < \infty$:

$$\int |f|^p \leq C_{N+1,0} \cdot \int (1+|x|)^{-(N+1)p} < \infty.$$

Since $C_0^\infty(\mathbb{R}^n)$ is dense in $L^p(\mathbb{R}^n)$ and its a subspace of $S(\mathbb{R}^n)$, finally it is proven that $\mathcal{S}(\mathbb{R}^n)$ is dense in $L^p(\mathbb{R}^n)$. It may not be entirely clear, however, why this integral is finite. This is explained in the following remark. □

Observation 5. *It needs to be justified why* $\int (1+|x|)^{-(N+1)p} < \infty$. *Indeed, notice that since one is integrating in* $\mathbb{R}^n$*, by symmetry, it can be seen that this integral coincides with the following integral:*

$$\int_{|x|\geq 1} \frac{1}{|x|^{p(N+1)}} dx.$$

In polar coordinates, one has $x = r\omega$ *and* $dx = r^{n-1} dr d\omega$. *Hence, the integral in polar coordinates takes the form:*

$$\int_1^\infty \int_{S^{n-1}} \frac{1}{|r\omega|^{p(N+1)}} r^{n-1} d\omega dr.$$

Since $|\omega| = 1$*, the integral in* ω *is a constant and, hence it can be written as:*

$$C \cdot \int_1^\infty \frac{1}{r^{p(N+1)-n+1}} dr,$$

which its known to converge if $p(N+1) - n + 1 > 1$, i.e., $N > (n/p) - 1$. As N is arbitrary, it suffices to choose its value properly and this integral will converge.

Now, it is important to define the convergence in this new space.

Definition 31. *A sequence of test functions in $\mathcal{S}$, $\{\varphi_k(x)\}_{k\in\mathbb{Z}_+^*}$, is said to converge to $\varphi(x)$ if and only if the functions φ_k and all their derivatives converge to φ and the corresponding derivatives of φ uniformly with respect to x in every bounded region R in $\mathbb{R}^n$. This sometimes is denoted as*

$$\varphi_k \xrightarrow{\mathcal{S}} \varphi.$$

The above definition implies that the constants $C_{\alpha\beta}$ that occur in Definition 30 can be chosen independently of x such that

$$|x^\alpha \partial^\beta \varphi(x)| < C_{\alpha\beta}$$

for all values of k. It is not very hard to show that $\varphi \in \mathcal{S}$ and, hence, $\mathcal{S}$ is closed with respect to this convergence.

Observation 6. *Another way to see that the space $\mathcal{D}$ is dense in $\mathcal{S}$ is to take an arbitrary C^∞ function $\beta(x)$ that is identically 1 for $|x| \leq 1$ and is identically 0 for $|x| \geq 2$. This β function can be constructed with the help of bump functions and its left to think about it. When $\varphi(x) \in \mathcal{S}$, the test functions*

$$\varphi_k(x) = \beta\left(\frac{x}{k}\right)\varphi(x) \qquad (k \in \mathbb{Z}_+^*)$$

are test functions belonging to $\mathcal{D}$ and such that the sequence $\{\varphi_k(x)\}$ converges to $\varphi(x)$ in the sense of $\mathcal{S}$, i.e., $\varphi_k \xrightarrow{\mathcal{S}} \varphi$.

The intuition behind the tempered distributions that will be soon defined are the same as the intuition of the usual distributions as before, *i.e.*, they can sometimes be seen as "generalized functions", but with the remark that this space will be more restricted as explained in the introduction. The Delta distribution is also an example of tempered distribution and its intuition here is the same.

Definition 32. *A* ***Tempered Distribution*** *(or sometimes called* ***Distribution of Slow Growth****) its a continuous linear functional in $\mathcal{S}(\mathbb{K}^n)$. The set of all tempered distributions is denoted by $\mathcal{S}'(\mathbb{K}^n)$ and is sometimes called the* ***Set of Schwartz Distributions****.*

The following lemma will be used to give another characterization of the convergence in the space $\mathcal{S}(\mathbb{K}^n)$.

Lemma 5. *Consider a linear functional $T : A \to \mathbb{C}$, where A is a vector space. Since $\mathbb{C}$ is also a vector space (over itself), then, for T to be continuous, it suffices that T is continuous at $0 \in A$.*

Proof. Assuming that T is continuous at 0, the goal is to show that T is continuous at every point. In fact, let $v, u \in A$. Given any $\varepsilon > 0$, one should find $\delta > 0$ such that:

$$|u - v| < \delta \Rightarrow |T(u) - T(v)| < \varepsilon.$$

Notice that, $|T(u) - T(v)| = |T(u - v) - 0| = |T(u - v) - T(0)|$, and $|u - v| = |(u - v) - 0|$. All those steps before are justified simply by the linearity of T and the fact that A and $\mathbb{C}$ are vector spaces. Now with that being said, notice that analysing the continuity in v, with $u = v + h$ such that $|h| < \varepsilon$, is equivalent to analysing the continuity at 0. Hence, it suffices to prove the continuity at 0 to assure that the described linear functional is continuous at every point. $\square$

A sequence $\varphi_k \xrightarrow{\mathcal{S}} \varphi$ with $\varphi \equiv 0$ is called a **null sequence**. Hence, by Lemma 5, another characterization of tempered distributions can be stated by saying that $T \in \mathcal{S}$ if and only if

1. $\langle T, c_1\varphi_1 + c_2\varphi_2\rangle = c_1\langle T, \varphi_1\rangle + c_2\langle T, \varphi_2\rangle$,

2. $\lim\limits_{k \longrightarrow \infty} \langle T, \varphi_k\rangle = 0$ for every null sequence $\{\varphi_k\}_{k \in \mathbb{Z}_+^*}$ with $\varphi_k \in \mathcal{S}$ for every k.

Since by definition it can be seen from the definitions of $\mathcal{D}$-convergence and §-convergence that a sequence $\{\varphi_k\}_{k\in\mathbb{Z}_+^*}$ converging to a function φ in the $\mathcal{D}$-convergence also converges to φ in the $\mathcal{S}$-convergence. Moreover, since every linear functional on $\mathcal{S}$ is also a linear functional on $\mathcal{D}$, then $\mathcal{S}' \subseteq \mathcal{D}$.

The good thing is that the vast majority of distributions discussed in the previous chapter are also tempered distributions.

Example 20. *As an example of a distribution $T \in \mathcal{D}$ such that $T \notin \mathcal{S}$ (i.e. a distribution that is not a tempered distribution) is given by the locally integrable function $f(x) = e^{x^2} \in \mathcal{D}'$ (again here is an abuse of notation, since the distribution itself is T_f). In fact $T \in \mathcal{D}$ but $T_f \notin \mathcal{S}$ as the reader can verify.*

The role played by the locally integrable functions $f \in L^1_{\text{Loc}}(\mathbb{K}^n)$ in $\mathcal{D}'$ completely analogous to the role played by functions of slow growth $f \in \mathcal{S}$ in the space $\mathcal{S}'$. This result is given in the following theorem. However, first, just to clarify, the concept of slow growth is used to classify functions in general, not only test functions and is formalized in the next definition:

Definition 33. *A function f, $x \longmapsto f(x)$, defined for $x \in \mathbb{R}^n$ is called a* ***Function Of Slow Growth*** *if f, and all of its derivatives, grows at infinity more slowly than some polynomial. That means that there exists constants C, k and A such that:*

$$|\partial^\alpha f(x)| \leq C|x|^k, \qquad |x| > A.$$

Now, the theorem:

Theorem 13. *Every function $f(x)$ of slow growth induces a distribution through the formula:*

$$\langle f, \varphi \rangle = \int f\varphi \, d\lambda, \qquad \varphi \in \mathcal{S},$$

where λ denotes the Lebesgue measure.

Proof. This operation $\langle f, \cdot \rangle$ is clearly a linear functional by the definition of integral. In order to prove the continuity, one should show that if $\{\varphi_k\}$ is a null-sequence in $\mathcal{S}$, then $\langle f, \varphi_k \rangle \longrightarrow 0$ as $k \longrightarrow \infty$. Indeed, for each k:

$$\int f(x)\varphi_k(x)\,dx = \int \frac{f(x)}{(1+|x|^2)}\left[(1+|x|^2)^\ell \varphi_k(x)\right]\,dx,$$

where $\ell \in \mathbb{Z}_+$. For a sufficiently large ℓ the factor

$$\frac{f(x)}{(1+|x|^2)^\ell}$$

is absolutely integrable, and hence:

$$\left|\int f(x)\varphi_k(x)\,dx\right| \le \sup(|(1+|x|^2)^\ell \varphi_k(x)|) \int \left|\frac{f(x)}{(1+|x|^2)^\ell}\right|\,dx.$$

Notice that the right hand side of the equation above goes to 0 as $\varphi_k \longrightarrow 0$. Therefore, $\langle f, \varphi_k \rangle \longrightarrow 0$ for a null sequence $\{\varphi_k\}$, which proves the continuity. □

An important fact is that $\mathcal{S}'$ contain some locally integrable functions (actually again the induced distributions) that are not of slow growth:

Example 21. *Consider the function* $f(x) = [\cos(e^x)]' = -e^x \sin(e^x)$. *Then f is clearly a function that is not of slow growth. Nonetheless,* $T_f \in \mathcal{S}'$ *since:*

$$\langle (\cos(e^x))', \varphi \rangle = -\int \cos(e^x)\varphi'(x)\,dx, \qquad \varphi \in \mathcal{S}.$$

The same is done to define convergence in $\mathcal{S}'$ as was to define in $\mathcal{D}'$, *i.e.*, as a weak convergence (see [3], [13]). The definition is given bellow:

Definition 34. *A sequence $\{T_k\}$ of distributions in $\mathcal{S}'$ is said to converge to $T \in \mathcal{S}'$ (a convergence known as* ***$\mathcal{S}'$-convergence****), and denoted by*

$$T_k \xrightarrow[k \longrightarrow \infty]{\mathcal{S}'} T,$$

if, for every test function $\varphi \in \mathcal{S}$, $\langle T_k, \varphi \rangle \longrightarrow \langle T, \varphi \rangle$ as $k \longrightarrow \infty$.

From the above definition and the fact that $\mathcal{S}' \subseteq \mathcal{D}'$, it follows that a sequence of tempered distributions $\{T_k\}$ converging to a distribution $T \in \mathcal{S}'$ converges also in $\mathcal{D}'$ to T, *i.e.*, according to the $\mathcal{D}'$ convergence.

These last results are summarized in the next theorem:

Theorem 14. *The following relations of set inclusion are valid: $\mathcal{D} \subseteq \mathcal{S}$ and $\mathcal{S}' \subseteq \mathcal{D}'$. Moreover, $\mathcal{D}$-convergence implies $\mathcal{S}$-convergence and $\mathcal{S}'$- convergence (weak convergence) implies $\mathcal{D}'$-convergence (weak convergence).*

Proof. It is just a sum of a few last results already proven. □

Just as before in the context of test functions, the good thing is that the majority of the distributions exposed here are in $\mathcal{S}'$. Furthermore, many of the operations defined for distributions in $\mathcal{D}'$ remain valid in $\mathcal{S}'$ since $\mathcal{S}'$ is a subspace of $\mathcal{D}'$. However, the problem is that the results of some operations for a tempered distribution may not result in a tempered distribution, and, if the results of some operation for tempered distribution produces a tempered distribution, then its said that $\mathcal{S}'$ is closed under such an operation.

Example 22. *Some operations that remain valid and for which $\mathcal{S}'$ is closed under are:*

1. *Addition.*

2. *Scalar Multiplication.*

3. *Linear change of variables.*

4. *Product of a distribution by a test function.*

5. *Differentiation.*

Example 23. *To provide an operation such that $\mathcal{S}'$ is not closed under, consider the multiplication by a smooth function. Consider the impulse train, which a countable summation of Delta distributions. Write $T(x)$ and interpret T as a function (like in the introduction of the chapter about distributions):*

$$T(x) = \sum_{k=1}^{\infty} \delta(x-k).$$

Then, $T \in \mathcal{S}'$. However, $e^{x^2}T(x)$ is not in $\mathcal{S}$, since, for $\varphi(x) = e^{-x^2} \in \mathcal{S}$, then $\langle \varphi(x), e^{x^2}T(x)\rangle = 1+1+\cdots+1+\cdots$, which is obviously divergent. Yet, if one takes $\varphi \in \mathcal{D}$, then

$$\langle \varphi(x), e^{x^2}T(x)\rangle = \sum_{k=1}^{\infty} e^{k^2}\varphi(k)$$

is a sum with only a finite number of nonzero terms, and hence, is convergent.

Observation 7. *In the next theorem, the concept of convergence in $\mathcal{S}(\mathbb{R}^n)$ will be used. In particular, again the convergence to zero will be used. Remember that in the general case, given a sequence of functions ϕ_n em $S(\mathbb{R}^n)$, the sequence is said to converge in $\mathcal{S}$ to $\phi \in S(\mathbb{R}^n)$ if $||\phi - \phi_n||_{\alpha,\beta} \longrightarrow 0$ as $n \longrightarrow \infty$.*

In particular, notice that if (ϕ_n) converges to 0 in $\mathcal{S}$, then (ϕ_n) must converge to 0 pointwise. At first, one could think that (ϕ_n) converges to a constant function. However, since $(\phi_n) \in L^1(\mathbb{R}^n)$ it must not be constant, as it should approach 0 going to infinity.

Theorem 15. *Functions $f \in L^p(\mathbb{R}^n)$ define tempered distributions.*

Proof. Given $f \in L^p(\mathbb{R}^n)$, consider the following functional: $T_f : S(\mathbb{R}^n) \to \mathbb{C}$ given by $T_f(\phi(x)) = \int_{\mathbb{R}^n} f(x)\phi(x)dx$. Obviously, it is a linear functional by the linearity of the integral. Thus it needs to be proven that is continuous. In order to do that, it suffices to prove that it is continuous at 0, since $S(\mathbb{R}^n)$ and $\mathbb{C}$ are vector spaces. Hence, notice that if $(\phi_n(x)) \longrightarrow 0$, *i.e.*, (ϕ_n) converges uniformly to the identically zero function, then every derivative of ϕ_n goes to zero. Thus $||\phi_n||_{\alpha,\beta} \longrightarrow 0$. It needs to be proven that $T_f(\phi(x)) \longrightarrow 0$.

However, this is an immediate consequence of Holder's Inequality. Let q be such that $1/p+1/q = 1$, then $\phi_n \in L^q$ (its already seen that $S(\mathbb{R}^n) \subset L^q$). Moreover, it is already known that $\phi_n(x) \longrightarrow 0$ and hence $||\phi_n||_q \longrightarrow 0$. Thus:

$$\left|\int_{\mathbb{R}^n} f(x)\phi_n(x)dx\right| \leq \int_{\mathbb{R}^n} |f(x)\phi_n(x)|dx \leq ||f||_p \cdot ||\phi_n||_q \longrightarrow 0.$$

□

In order to make the notation shorter and more compact, just as was done in the case of $\mathcal{D}$, it is useful to identify as f the tempered distribution induced by $f \in L^p$, and write as $\langle f,\phi\rangle = \int_{\mathbb{R}^n} f(x)\phi(x)dx$.

3.2 Fourier Transform

In this second section, the Fourier Transform will be studied and its particularities when restricted to certain sets such as L^1, L^2 and $\mathcal{S}$. This study is extremely important to formalized this tool that is so used in a vast number of areas that concern mathematics, physics and engineering.

Firstly, the Fourier Transform will be studied when applied to functions in $L^1(\mathbb{R}^n)$. The constant factor that appears in the transform may be different and it depends of the area of study and its applications. In this section, it is chosen to be $\left(\frac{1}{2\pi}\right)^{\frac{n}{2}}$.

Definition 35. *Let $f \in L^1(\mathbb{R}^n)$. The **Fourier Transform** of f is a function $\hat{f} = \mathcal{F}[f] : \mathbb{R}^n \longrightarrow \mathbb{C}$ given by:*

$$\mathcal{F}[f](\xi) = \hat{f}(\xi) = \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \int_{\mathbb{R}^n} f(x) \cdot e^{-i\xi\cdot x}dx,$$

where $\xi \cdot x$ stands for the canonical dot product in $\mathbb{R}^n$.

Theorem 16. *If $f \in L^1(\mathbb{R}^n)$, then $\hat{f}(\xi)$ is bounded, continuous and satisfies:*

$$||\hat{f}||_\infty \leq \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} ||f||_1.$$

Proof. Notice that for all $\xi \in \mathbb{R}^n$:

$$\begin{aligned}
|\hat{f}(\xi)| &= \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \left|\int_{\mathbb{R}^n} f(x) \cdot e^{-i\xi \cdot x} dx\right| \\
&\leq \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \int_{\mathbb{R}^n} |f(x) \cdot e^{-i\xi \cdot x}| dx \\
&= \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \int_{\mathbb{R}^n} |f(x)| dx \\
&= \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} ||f||_1
\end{aligned}$$

Hence, since this is valid for all ξ, then $||\hat{f}||_\infty \leq \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} ||f||_1$.

In order to prove the continuity, one should show that if $\xi_n \to \xi$ then, $\hat{f}(\xi_n) \to \hat{f}(\xi)$. Indeed, if $\xi_n \to \xi$, then it suffices to take the limit in

$$|\hat{f}(\xi) - \hat{f}(\xi_n)| = \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \left|\int_{\mathbb{R}^n} f(x) \cdot (e^{-i\xi \cdot x} - e^{-i\xi_n \cdot x}) dx\right|.$$

Since $\lim_{n\to\infty}(e^{-i\xi \cdot x} - e^{-i\xi_n \cdot x}) = 0$, and $f(x)$ is bounded almost everywhere, then $\hat{f}(\xi_n) \to \hat{f}(\xi)$. ◻

Up to now, the Fourier Transform was defined for functions in $L^1(\mathbb{R}^n)$. However, it is easy to see that this space is not invariant with respect to this transform, as shown in the next example

Example 24. *Consider the characteristic function $f = \mathbb{1}_{[-1,1]}$ of the set $[-1, 1]$. This function is obviously integrable, and, therefore, belongs to $L^1(\mathbb{R})$. However, it is easy to verify that its Fourier Transform is* *given by:*

$$\hat{f}(\xi) = \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \cdot \frac{\sin(\xi)}{\xi}$$

One important remark is that at $\xi = 0$ one has $\hat{f}(0) = 1$. Hence, $\hat{f}$ is well defined, and it is already known that this transformed function does not belong to $L^1(\mathbb{R})$.

Now it is important to see some of the properties of the Fourier Transform when restricted to $\mathcal{S}(\mathbb{R}^n)$. One of the main goals is to show that this space is invariant with respect to the Fourier Transform. Moreover a inverse transform can be determined and this will make the Fourier Transform a linear automorphism when restricted to this space.

Before starting to verify these claims, it is important to establish two properties of the Fourier Transform in the space $\mathcal{S}(\mathbb{R}^n)$.

Lemma 6. *Let $f \in S(\mathbb{R}^n)$. Then, the following properties hold:*

- $(\partial^\beta \hat{f})(\xi) = (-i)^{|\beta|}\widehat{(x^\beta f)}$
- $\xi^\alpha \hat{f}(\xi) = (-i)^{|\alpha|}\widehat{(\partial^\alpha f)}$

Proof. The proof will be made specifically to the case $n = 1$. For the general case, the proof is longer, but the idea is the same. To prove the first property, it is enough to differentiate the formula for the transform with respect to ξ, and the result is immediate. During this proof, the normalizing constant will be leaved out to make the calculations cleaner.

For the second part, integration by parts will be used:

$$\int f'(x)e^{-ix\cdot\xi}dx = f(x)e^{-ix\cdot\xi}\Big|_{-\infty}^{+\infty} + i\xi \cdot \int f(x)e^{-ix\cdot\xi}dx.$$

Since f belongs to $\mathcal{S}$, it is integrable, and surely goes to 0 as $|x| \longrightarrow \infty$. As the exponential $e^{-ix\cdot\xi}$ is bounded, the term $f(x)e^{-ix\cdot\xi}|_{-\infty}^{+\infty}$ is equal to 0. Hence, it follows that $\xi\hat{f}(\xi) = (-i)\widehat{f'}$. By induction in the number of derivatives, the claim follows. $\square$

With these results in hand, it is possible to begin to describe the properties of the Fourier Transform in the space $\mathcal{S}(\mathbb{R}^n)$.

Theorem 17. *The space $S(\mathbb{R}^n)$ is invariant under the Fourier Transform, i.e., if $f \in S(\mathbb{R}^n)$, then, $\hat{f}(\xi) \in S(\mathbb{R}^n)$.*

Proof. Multiplying the two properties established in Lemma 6 and taking the absolute value:

$$|\xi^{\alpha}(\partial^{\beta}\hat{f})(\xi)| \cdot |\hat{f}(\xi)| = |\widehat{(x^{\beta}f)}| \cdot |\widehat{(\partial^{\alpha}f)}|.$$

Looking to the right hand side of the equation, notice that, from the relation $|\partial^{\alpha}f(x)| \leq C_{N,\alpha} \cdot (1 + |x|)^{-N}$, where it is possible to take α as the zero vector (*i.e.*, $|f(x)|$), and N as any positive integer, it is possible to conclude that, independently from the chosen β, one has that $\partial^{\alpha}f(x)$ and $x^{\beta}f(x)$ both belong to $L^1(\mathbb{R}^n)$. Hence, it is possible to apply Theorem 16, and it leads to the following inequality:

$$|\xi^{\alpha}(\partial^{\beta}\hat{f})(\xi)| \cdot |\hat{f}(\xi)| \leq \left(\frac{1}{2\pi}\right)^n ||\partial^{\alpha}f(x)||_1 \cdot ||x^{\beta}f(x)||_1 < \infty.$$

Since this holds for all $\xi \in \mathbb{R}^n$, it follows that $||\xi^{\alpha}(\partial^{\beta}\hat{f})(\xi)||_{\alpha,\beta} \leq \frac{1}{||\hat{f}(\xi)||_{\infty}} \cdot \left(\frac{1}{2\pi}\right)^n ||\partial^{\alpha}f(x)||_1 \cdot ||x^{\beta}f(x)||_1 < \infty$. Therefore, $\hat{f}(\xi) \in S(\mathbb{R}^n)$.

$\square$

Now that it is already proven that $\mathcal{S}(\mathbb{R}^n)$ is invariant with respect to the Fourier Transform, it is important to seek for a possible inverse transform when this transform is restricted to $\mathcal{S}(\mathbb{R}^n)$. The goal with this is to finally conclude that indeed this mapping is a isomorphism.

The next theorem provides exactly this result and it is known as the **Fourier's Inversion Theorem** for functions restricted to the space $\mathcal{S}$. The proof for this theorem will be done later, after the subject of convolution is presented.

Theorem 18 (Fourier's Inversion). *Let $f \in \mathcal{S}(\mathbb{R}^n)$ and $\hat{f}(\xi)$ be its Fourier Transform. Then,*

$$f(x) = \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \int_{\mathbb{R}^n} \hat{f}(\xi) e^{i\xi \cdot x} d\xi \ \forall x \in \mathbb{R}^n.$$

Proof. Will be done later, after presenting the convolution. $\square$

It is very interesting to notice the resemblance between the formulas for the Fourier Transform and its Inverse Transform. The inverse Fourier Transform of a function $f(\xi)$ will be denoted by $\mathcal{F}^{-1}[f](x) = \check{f}(x)$. Observe that $\check{f}(x) = \hat{f}(-x)$. Moreover, notice that if one has $\phi \in S(\mathbb{R}^n)$, then $(\hat{\phi})^{\vee} = \phi = (\check{\phi})^{\wedge}$. This shows that the Fourier Transform restricted to the Schwartz's Space is an isomorphism, since one should be aware that it is a bijection, it is continuous and its inverse is also continuous.

This concludes the subject for the Fourier Transform in L^1 and in $\mathcal{S}$. Now, the main focus will be to develop the theory for the Fourier Transform in L^2.

The idea to extend the Fourier Transform to L^2 is to use the density of $\mathcal{S}(\mathbb{R}^n)$ in $L^2(\mathbb{R}^n)$. Therefore, it becomes reasonably clear what one should try to do. The reasoning will be to look to sequences of functions in $\mathcal{S}$ converging to functions in L^2, and then analyse how the sequence of transforms behaves. Before beginning theses steps it is important to state an useful lemma which will help developing this theory.

Lemma 7 (Parseval's Identity). *Let $\phi, \psi \in \mathcal{S}(\mathbb{R}^n)$. Then,*

$$(\phi|\psi) = \int \phi(x)\overline{\psi}(x)dx = \int \hat{\phi}(\xi)\overline{\hat{\psi}}(\xi)d\xi = (\hat{\phi}|\hat{\psi}).$$

Proof. In this proof the constant will be omitted to simplify the calculations. By the Inversion Formula:

$$\begin{aligned}
(\phi|\psi) &= \int_{\mathbb{R}^n} \phi(x) \overline{\int_{\mathbb{R}^n} \hat{\psi}(\xi) e^{i\xi x} d\xi} dx \\
&= \int_{\mathbb{R}^n} \overline{\hat{\psi}(\xi)} \int_{\mathbb{R}^n} \phi(x) e^{-i\xi x} dx d\xi \\
&= \int_{\mathbb{R}^n} \hat{\phi}(\xi) \overline{\hat{\psi}(\xi)} d\xi \\
&= (\hat{\phi}|\hat{\psi}).
\end{aligned}$$

□

Notice that, if one takes $\phi = \psi$, then $(\phi|\phi) = ||\phi||_2^2$. Hence, by Parseval's Identity, it is possible to conclude an interesting property about the Fourier Transform and its relation with the norm in L^2. Indeed, observe that from the Parseval's Identity one has $||\phi||_2 = ||\hat{\phi}||_2$, *i.e.*, in the norm of L^2, the Fourier Transform is an isometry. Furthermore, notice that if a sequence of functions in $\mathcal{S}$ is convergent in L^2, then the sequence of transforms also converges in L^2.

It is already known that $\mathcal{S}(\mathbb{R}^n)$ is dense in $L^2(\mathbb{R}^n)$. Therefore, given $f \in L^2(\mathbb{R}^n)$, then there exists $(\phi_n) \subset \mathcal{S}(\mathbb{R}^n)$ such that $||f - \phi_n||_2 \to 0$ if $n \to \infty$. By Parseval's Identity, since (ϕ_n) converges in $L^2(\mathbb{R}^n)$, then $(\hat{\phi_n})$ also converges in $L^2(\mathbb{R}^n)$, *i.e.*, there exists $\varphi = \lim_{k\to\infty} \hat{\phi_k}$ in $L^2(\mathbb{R}^n)$. It is tempting to define this limit as $\hat{f}$. However, extra care must be taken. Since convergence in L^2 considers almost everywhere equivalences, it needs to be verified that φ does not depend on the convergence sequence that is chosen. Fortunately, this is easy to prove.

Lemma 8. *Let* $(\phi_n), (\psi_n) \subseteq \mathcal{S}(\mathbb{R}^n)$ *sequences converging to* f *in* $L^2(\mathbb{R}^n)$. *Let* $\Phi = \lim \widehat{\phi_n}$ *and* $\Psi = \lim \widehat{\psi_n}$, *then* $||\Phi - \Psi||_2 = 0$.

Proof. Due to the convergence hypothesis, it can be seen that, given $\varepsilon > 0$, there exists $n_0 \in \mathbb{Z}_+$ such that for all $n > n_0$:

- $||\Phi - \widehat{\phi_n}||_2 < \frac{\varepsilon}{3}$.
- $||\Psi - \widehat{\psi_n}||_2 < \frac{\varepsilon}{3}$.
- $||\widehat{\psi_n} - \widehat{\phi_n}||_2 < \frac{\varepsilon}{3}$.

Hence, using the triangle inequality:

$$\begin{aligned}||\Phi - \Psi||_2 &\leq ||\Phi - \widehat{\phi_n}||_2 + ||\widehat{\psi_n} - \widehat{\phi_n}||_2 + ||\Psi - \widehat{\psi_n}||_2 \\ &< \varepsilon.\end{aligned}$$

Thus, it is possible to conclude that, in L^2, the convergence of the transforms of a sequence that converges to $f \in L^2$ is also convergent and is independent of the chosen sequence to approximate. $\square$

Another way to prove the result above would be to consider it as a very particular case of the following theorem:

Theorem 19 (Bounded Linear Transformation (B.L.T.)). *Suppose G is a bounded linear transformation from a normed linear space $(V_1, \|\cdot\|_{V_1})$ to a complete normed linear space $(V_2, \|\cdot\|_{V_2})$. Then G can be uniquely extended to a bounded linear transformation $\tilde{G}$ (with the same bound), from the completion of V_1 to $(V_2, \|\cdot\|_{V_2})$.*

Proof. See [15]. $\square$

With these facts in hand, the definition of the Fourier Transform in L^2 becomes extremely natural. Furthermore, observe that the isometry property seen in the norm L^2 will be valid for all functions in this case and not only to functions belonging to $\mathcal{S} \cap L^2$.

Definition 36. *Let $f \in L^2(\mathbb{R}^n)$. Then its Fourier Transform $\hat{f}$ is defined as:*

$$\hat{f} = \lim_{n \to \infty} \widehat{\phi_n},$$

where (ϕ_n) *is any sequence in* $\mathcal{S}(\mathbb{R}^n)$ *converging to* f *in* $L^2(\mathbb{R}^n)$*.*

In order to define the inverse transform, it suffices to proceed in an entirely analogous method, *i.e.*, take a sequence in $\mathcal{S}$ converging to a function $g \in L^2$, and then define its Inverse Fourier Transform as the limit of the inverse transforms of the functions in this sequence. Therefore, the Fourier Transform in L^2, aside from being an isometry, also maintain the property of being an isomorphism.

Finally, the last space for which the Fourier Transform will be defined here is the space $\mathcal{S}'$, *i.e.*, the space of the continuous linear functionals in $\mathcal{S}$. To motivate this definition, consider the following example. Let $f \in L^1(\mathbb{R}^n)$. Remembering how f defines a tempered distribution, consider the following calculation:

$$\begin{aligned}
\langle \hat{f}, \phi \rangle &= \int \hat{f}(\xi)\phi(\xi)d\xi \\
&= \int \phi(\xi)\left(\int f(x)e^{-i\xi\cdot x}dx\right)d\xi \\
&= \int f(x)\left(\int \phi(\xi)e^{-i\xi\cdot x}d\xi\right)dx \\
&= \int f(x)\hat{\phi}(x)dx = \langle f, \hat{\phi} \rangle.
\end{aligned}$$

With this example in hand, it can be seen that one has two tempered distributions (one defined by f and the other by $\hat{f}$) and such that the relation between them is that, applying $\phi \in S(\mathbb{R}^n)$ in one, this leads to the same result if one applies $\hat{\phi}$ in the other. This motivates a natural definition for the Fourier Transform for tempered distributions, as can be seen next.

Definition 37. *Let* $T \in \mathcal{S}'(\mathbb{R}^n)$*. The Fourier Transform of* T*, denoted by* $\hat{T}$*, is also a tempered distribution such that:*

$$\langle \hat{T}, \phi \rangle = \langle T, \hat{\phi} \rangle \quad \forall \phi \in \mathcal{S}(\mathbb{R}^n).$$

Example 25. *The best way to understand what this definition really means is by giving an example. In* *that sense, the classical example of the Fourier Transform of the Dirac's Delta suits very well. This example is extremely important, since this distribution is present in numerous study areas such as the modeling of systems in engineering. This last observation about modeling of system is due to, when an extremely strong action occurs almost instantly, the Dirac's Delta may be a good easy to describe this phenomenon.*

Applying the definition of the Dirac's Delta centered at 0*:* $\langle \delta, \hat{\phi} \rangle = \hat{\phi}(0)$. *By the Inversion Formula, it is known that*

$$\hat{\phi}(0) = \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \int_{\mathbb{R}^n} \phi(x)dx = \left\langle \left(\frac{1}{2\pi}\right)^{\frac{n}{2}}, \phi \right\rangle.$$

Hence, $\langle \hat{\delta}, \phi \rangle = \langle \left(\frac{1}{2\pi}\right)^{\frac{n}{2}}, \phi \rangle$. *Therefore one can conclude that the Fourier Transform of the Dirac's Delta is represented by a constant. In this case, accordingly to the definition of Fourier Transform in this chapter, that constant is* $\left(\frac{1}{2\pi}\right)^{\frac{n}{2}}$.

Proposition 11. *The Fourier Transform defined on* $\mathcal{S}'$ *also maps to elements in* $\mathcal{S}'$ *and, furthermore, it is continuous.*

Proof. Let $T \in \mathcal{S}'$. The goal is to show that indeed $\hat{T}$ belongs to $\mathcal{S}'$. By the definition of $\hat{T}$, one has that, for all $\phi \in S$, is true that $|\langle \hat{T}, \phi \rangle| = |\langle T, \hat{\phi} \rangle|$. Due to the characterization of linear operators in $\mathcal{S}$, the following inequality holds:

$$|\langle \hat{T}, \phi \rangle| \leq C \cdot \sum_{|\alpha|,|\beta| \leq N} |\xi^\alpha \partial^\beta \hat{\phi}(\xi)|.$$

Now, by the properties of the Fourier Transform presented in Lemma **??**, $|\xi^\alpha \partial^\beta \hat{\phi}(\xi)| = |\widehat{\partial^\alpha (x^\beta f(x))}| \leq ||\partial^\alpha (x^\beta f(x))||_1$. Since the norm L^1 of functions in $\mathcal{S}$ can be bounded using multi-indices α, β, there must exist an N' such that:

$$|\langle \hat{T}, \phi \rangle| \leq C \cdot \sum_{|\alpha|,|\beta| \leq N} |\xi^\alpha \partial^\beta \hat{\phi}(\xi)| \leq C' \cdot \sum_{|\alpha|,|\beta| \leq N'} |x^\alpha \partial^\beta \phi(x)|.$$

Therefore, by the description of linear operators in $\mathcal{S}'$, one has that, indeed, $\hat{T} \in \mathcal{S}'$.

In order to prove the continuity, the definition of $\mathcal{S}'$-convergence mus be remembered. This is done by sequential continuity, *i.e.*, $T_n \longrightarrow T$ if, for all $\phi \in \mathcal{S}$, then $\langle T_n, \phi \rangle \longrightarrow \langle T, \phi \rangle$. That being said, it is easy to verify the continuity of the Fourier Transform in $\mathcal{S}'$. Indeed, let $T_n \to T$, then, for all $\varepsilon > 0$, there exists $n_0 \in \mathbb{Z}_+$ such that, $\forall n > n_0$ $|\langle T_n, \phi \rangle - \langle T, \phi \rangle| < \varepsilon$, for any $\phi \in \mathcal{S}$. Hence, applying the definition of the transform:

$$\begin{aligned}
|\langle \widehat{T_n}, \phi \rangle - \langle \widehat{T}, \phi \rangle| &= |\langle \widehat{T_n - T}, \phi \rangle| \\
&= |\langle T_n - T, \widehat{\phi} \rangle| \\
&= |\langle T_n, \widehat{\phi} \rangle - \langle T, \widehat{\phi} \rangle| \to 0.
\end{aligned}$$

□

Therefore, applying the process in an analogous way, it may be verified that the inverse of the transform is also continuous, in a way that the Fourier Transform in $\mathcal{S}'$ is an isomorphism. This is stated in the next theorem.

Theorem 20. *Both the Fourier Transform and the Inverse Fourier Transform defined on $\mathcal{S}'$ maps to elements in $\mathcal{S}'$ and, furthermore, both are continuous. Therefore, the Fourier Transform is an isomorphism in $\mathcal{S}'$.*

Proof. This is just a synthesis of the previous results. □

3.3 Convolution

The convolution here is defined the same way as it was in the previous chapter. One of the main goals is, as said in the last section, to prove the Fourier Inversion Formula.

Just to recapitulate, for two functions f and g with $\mathbb{R}^n$ as domain, the convolution between f and g is defined as

$$(f * g)(x) = \int_{\mathbb{R}^n} f(x-y)g(y)dy,$$

whenever this integral makes sense.

Now, some important properties of the convolution operation will be stated. In particular, a fundamental property in the theory of differential equations will be proved, the *Smoothing Property*. In general lines, the convolution operation can return a function that is more "regular". For example, if both f and g and their respective derivatives are in L^1, then the second derivative of their convolution lies in L^1. It will be seen the case of the convolution of a smooth function with a function in L^1.

Proposition 12. *The following properties hold for the convolution:*

1. $(f * g)(x) = (g * f)(x)$.
2. *If f and $g \in L^1(\mathbb{R}^n)$, then* $||(f * g)||_1 \leq ||f||_1 \cdot ||g||_1$.
3. *Let f and g be compactly supported functions. Then $(f*g)$ has compact support and* $\mathrm{supp(f*g)} \subset \{\mathrm{w+y : w \in supp(f), y \in supp(g)}\}$.
4. *If $f \in C^k_c(\mathbb{R}^n)$ and $g \in L^1(\mathbb{R}^n)$ then,* $(f * g) \in C^k(\mathbb{R}^n)$.

Proof.

1. It suffices to make a chance of variables $z = x - y$. Indeed:

$$(f * g)(x) = \int_{\mathbb{R}^n} f(x-y)g(y)dy = -\int_{\mathbb{R}^n} f(z)g(x-z)(-1)dz = (g * f)(x).$$

2.

$$\begin{aligned}
||(f*g)||_1 &= \int_{\mathbb{R}^n}\left|\int_{\mathbb{R}^n} f(x-y)g(y)dy\right|dx \\
&\leq \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} |f(x-y)||g(y)|dxdy \\
&= ||f||_1 \cdot ||g||_1.
\end{aligned}$$

3. Notice that, under the conditions in this item $(f*g)(x) = \int_{\mathbb{R}^n} f(x-y)g(y)dy$ is not identically zero only if $y \in \text{supp}(g)$ e $x - y = w \in \text{supp}(f)$. Then, it is possible that $(f*g)(x) \neq 0$ if $x = w + y$, proving, therefore, that the support of the convolution is contained in the union of the supports of f and g.

4. This last property is basically application of the Lebesgue Dominated Convergence Theorem (see [2]). In fact, what should be proved is that, for functions satisfying the conditions enunciated, it holds that: $\partial_{x_1}(f*g)(x) = (\partial_{x_1}f*g)(x)$. It suffices to show that for only one derivative, since the general case follows from mathematical induction. In order to prove that, consider the following calculations:

$$\partial_{x_1}(f*g)(x) = \lim_{h\to 0}\int_{\mathbb{R}^n}\frac{f(x+he_1-y)-f(x-y)}{h}g(y)dy.$$

To switch the positions of the limit and the integral, it needs to be shown that the integrand is dominated by an integrable function. Indeed, since f has compact support, it is known that its derivative is limited with an upperbound M. Moreover, it is also known that, since g is integrable, the integrand is indeed dominated by $M|g(y0)|$ and, thus, integrable. Hence, by the Lebesgue Dominated Convergence Theorem, the integral and the limit can switch places and the desired result follows, $\partial_{x_1}(f*g)(x) = (\partial_{x_1}f*g)(x)$.

$\square$

In the first proposition, a very shallow characterization was established for the relationship between the spaces in which f and g lie and the space in which the convolution lies. In fact, a basic relation was shown, *i.e.*, the fact that if f and g both lie in L^1, then the convolution also lies in L^1. However there exists a much stronger result than that, known as the *Young's Inequality For The Convolution*. This result will be presented now, but is proof will only be done in Section A.1. The proof for this result uses the *Riesz-Thorin's Interpolation Theorem*, which will be the central subject of Section A.1. More about this can be seen in [6].

Proposition 13. *Let $f \in L^p$ and $g \in L^q$. Also, let r be such that $1 + \frac{1}{r} = \frac{1}{p} + \frac{1}{q}$, then, Young's Inequality holds: $||f * g||_r \leq ||f||_p \cdot ||g||_q$. In particular, if p and q are conjugate, i.e., $p^{-1} + q^{-1} = 1$ then $f * g$ is bounded (due to Young's Inequality), uniformly continuous and, if $1 < p, q < \infty$, then $f * g \in C_0(\mathbb{R}^n)$, i.e., decreases to zero approaching infinity.*

Proof. The proof of Young's Inequality will be done in Section A.1. If p or q are equal to 1, then the conjugate is equal to infinity, and r will also be infinity. The uniform continuity follows from the continuity of the translation in L^p (Lebesgue measure is invariant under translations). Since the same property will appear in the most interesting case, *i.e.*, when p and q are both different from 1 and infinity, this second part will be proved.

<u>Uniform continuity:</u> Let $f \in L^p$ and $g \in L^q$ with p and q conjugate. Then:

$$|f * g(x+h) - f * g(x)| \leq ||f(x+h-y) - f(x-y)||_p \cdot ||g||_q.$$

Since translation in L^p is continuous (it suffices to prove this for characteristic functions, then extend by linearity to simple functions and apply a density argument), then $|f * g(x+h) - f * g(x)| \longrightarrow 0$ if $h \longrightarrow 0$ and, thus, the result follows since this does not depend on x.

Decreasing at infinity: As seen before, the class of continuous functions with compact support is dense in L^p. Therefore, given $\varepsilon > 0$, choose $\tilde{f} \in C_c^\infty$ such that

$$||f - \tilde{f}||_p < \frac{\varepsilon}{2||g||_q},$$

and, in an analogous procedure, choose $\tilde{g} \in C_c^\infty$ such that

$$||g - \tilde{g}||_q < \frac{\varepsilon}{2||\tilde{f}||_p}.$$

Using the equality

$$f * g(x) = (f - \tilde{f}) * g(x) + \tilde{f} * (g - \tilde{g})(x) + \tilde{f} * \tilde{g}(x),$$

then using the Triangle and Young's Inequalities (r equal to infinity):

$$\begin{aligned} |f * g(x)| &\leq ||f - \tilde{f}||_p \cdot ||g||_q + ||\tilde{f}||_p \cdot ||g - \tilde{g}||_q + |\tilde{f} * \tilde{g}(x)| \\ &< \varepsilon + |\tilde{f} * \tilde{g}(x)|. \end{aligned}$$

Since $\tilde{f}, \tilde{g}$ are compactly supported, then Proposition 12 implies that the convolution is also compactly supported. Therefore, there exists a compact set K such that, outside K, $f * g(x)$ is less than ε, which characterizes the decrease to 0 at infinity. □

Now that the fundamental property of smoothing for the convolution has been presented, it is important to recall some facts about mollifiers also called approximations to the identity (see Chapter 2). Remember that the idea here is to find a sequence of functions with "good regularity" (in fact here C^∞) that converge to the original function f for the problem. The motivation here is to present a convolution with the Dirac's Delta. Indeed, considering the Delta as a singular measure, it means that $\int f(x-y)d\delta(y) = \int f(x-y)\delta(y)dy = f(x)$. The goal is to find a sequence of functions that approximate the Delta in terms of convolution.

It is important to also recall from Chapter 2 that a useful type of functions to be used in the identity approximations are the bump functions.

Example 26. *Let $\varphi \in L^1(\mathbb{R}^n)$. Define $\varphi_t(x) = \frac{1}{t^n}\varphi\left(\frac{x}{t}\right)$ for $t > 0$. Then,*

- *Assuming $y = \frac{x}{t}$:*

$$\begin{aligned}\int_{\mathbb{R}^n} \varphi_t(x)dx &= \int_{\mathbb{R}^n} \frac{1}{t^n}\varphi\left(\frac{x}{t}\right)dx \\ &= \int_{\mathbb{R}^n} \varphi(y)dy\end{aligned}$$

- *Notice that if $\varphi(x)$ has compact support, then $\varphi_t(x)$ also has compact support, and, as t decreases, this support also decreases.*

With that being said, the main theorem about identity approximations may be stated. This theorem has two versions, that will be called the "weak version" and the "strong version".

Theorem 21. *Let $\varphi \in L^1(\mathbb{R}^n)$ and $\int_{\mathbb{R}^n} \varphi(x)dx = 1$, then:*

1. ***Weak Version:** If $f \in L^p(\mathbb{R}^n)$ for $1 \leq p < \infty$, then $f * \varphi_t \to f$ in $L^p(\mathbb{R}^n)$ when $t \to 0$.*

2. ***Strong Version:** If there exists $C, \varepsilon > 0$ such that*

$$\varphi(x) \leq \frac{C}{(1+|x|)^{n+\varepsilon}},$$

then $f * \varphi_t \to f$ almost everywhere (a.e.).

Proof. 1. Taking $y = t\omega$, then:

$$\left(\int_{\mathbb{R}^n} |f(x) - f * \varphi_t(x)|^p \, dx\right)^{\frac{1}{p}} = \left(\int_{\mathbb{R}^n} \left|\int_{\mathbb{R}^n} (f(x) - f(x-y))\varphi_t(y)dy\right|^p dx\right)^{\frac{1}{p}}$$
$$= \left(\int_{\mathbb{R}^n} \left|\int_{\mathbb{R}^n} (f(x) - f(x-t\omega))\varphi(\omega)d\omega\right|^p dx\right)^{\frac{1}{p}}$$

and, using Minkowski's Ineuality,

$$\leq \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} |f(x) - f(x-t\omega)|^p dx\right)^{\frac{1}{p}} \varphi(\omega)d\omega$$
$$= ||f(x) - f(x-t\omega)||_p.$$

Recalling that the translations are continuous in L^p, taking $t \longrightarrow 0$ the result follows.

2. In order to prove the a.e. convergence, consider:

$$|f - f * \varphi_t| = \left|\int_{\mathbb{R}^n} (f(x) - f(x-t\omega))\varphi(\omega)d\omega\right|$$
$$\leq \int_{\mathbb{R}^n} |(f(x) - f(x-t\omega))\varphi(\omega)|d\omega$$
$$\leq \int_{\mathbb{R}^n} |(f(x) - f(x-t\omega))| \left|\frac{C}{(1+|x|)^{n+\varepsilon}}\right| d\omega.$$

Since $\dfrac{C}{(1+|x|)^{n+\varepsilon}}$ belongs to L^q with $q^{-1} + p^{-1} = 1$, Holder's Inequality implies that

$$|f - f * \varphi_t| \leq C' ||f(x) - f(x-t\omega)||_p,$$

and, using the continuity of translations in L^p the proof is done.

□

One interesting fact about mollifiers is that using them, it is possible to obtain a sequence of functions with "good regularity", since there exists sequences of C^∞ functions that converge to the desired function.

Finally, in this section of convolution, one important result still remains to be proved: the Fourier Inverse Transform Theorem. The proof was left to be done here because it becomes much simpler. During the proof, some basic properties and calculations concerning Fourier Transform will be used and are presented in the next proposition.

Proposition 14.

- *If, by the Fourier Transform,* $f(x) \longrightarrow \hat{f}(\xi)$, *then* $e^{iyx}f(x) \longrightarrow \hat{f}(\xi - y)$.
- *If* $y(x) = e^{\frac{-|x|^2}{2}}$, *then* $\hat{y} = y$ *and* $\int y(x)dx = (2\pi)^{\frac{n}{2}}$.

Proof. See [13] or another reference with Fourier Transforms Tables. □

With this properties in hand, and some important facts about the mollifiers revisited, it will be possible to prove the Fourier Inverse Transform Theorem wich will be restated here.

Theorem 22 (Fourier Inversion). *Let* $f(x)$ *and* $\hat{f}(\xi)$ *be functions in* $L^1(\mathbb{R}^n)$. *Then:*

$$f(x) = \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \int_{\mathbb{R}^n} \hat{f}(\xi) e^{ix\cdot\xi} d\xi.$$

Proof. First, define $\varphi(\xi) = e^{i\xi x} \cdot e^{-t^2\frac{|\xi|^2}{2}}$. Calculating the Fourier Transform of this function:

$$\hat{\varphi}(y) = \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \int_{\mathbb{R}^n} e^{i\xi x} \cdot e^{-t^2\frac{|\xi|^2}{2}} \cdot e^{-i\xi y} d\xi$$

and, making $t\xi = k$,

$$\begin{aligned}\hat{\varphi}(y) &= \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \left(\frac{1}{t^n}\right) \int_{\mathbb{R}^n} e^{ik\frac{x}{t}} \cdot e^{-\frac{|k|^2}{2}} \cdot e^{-ik\frac{y}{t}} dk \\ &= \left(\frac{1}{t^n}\right) e^{-\frac{|x-y|^2}{2t^2}}.\end{aligned}$$

The properties of Proposition 14 were used. Notice that the Fourier Transform of the function φ can be seen as a certain $g_t(x-y)$ in which $g(x) = e^{-\frac{|x|^2}{2}}$ in the sense that was defined for the approximations of the identity. Moreover, also by Proposition 14, it is known that $\displaystyle\int g_t(x-y)dy = \int g(y)dy = (2\pi)^{\frac{n}{2}}$.

It was seen in the study of the Fourier Transform that, for functions in L^1, the following identity holds: $\displaystyle\int f\hat{\varphi} = \int \hat{f}\varphi$. Using this and denoting by $h_t(x-y) = \dfrac{g_t(x-y)}{(2\pi)^{\frac{n}{2}}}$, this leads to:

$$\begin{aligned}\left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \int_{\mathbb{R}^n} e^{-t^2\frac{|\xi|^2}{2}} \hat{f}(\xi) e^{i\xi x} d\xi &= \int_{\mathbb{R}^n} h_t(x-y) f(y) dy \\ &= (h_t * f)(x).\end{aligned}$$

Applying the limit as $t \longrightarrow 0$, by the Lebesgue Dominated Convergence Theorem, (that can in fact be used here since all the functions are integrable), the left hand side goes to $\displaystyle\left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \int_{\mathbb{R}^n} \hat{f}(\xi) e^{ix\cdot\xi} d\xi$, and the right hand side to $f(x)$ by Theorem 21 (the results about mollifiers). Hence, the theorem is proved. $\square$

It's fine to celebrate success, but it's more important to heed the lessons of failure.

Bill Gates

4

Wavelets and Wavelet Transform

The goal of this chapter is to construct the theory of *wavelets* and the *Wavelet Transform*, comparing it with the Fourier Transform pointing out the advantages and the disadvantages of each theory. A relation with distribution theory will be established and, at the end, an application in signal processing will be presented concerning different frequency analysis methods applied to a chirp signal with high frequency interjections.

4.1 Introduction To Wavelets And The Short Time Fourier Transform

The usual Fourier Transform (FT) is defined for some "well behaved" functions. Here, the concern will be with maps $f : \mathbb{R} \to \mathbb{R}$. The FT can be defined (here the definition will be not normalized as it is more usual in the context of signal processing) as follows (see Chapter 3 for details in discussion):

$$\hat{f}(\omega) = \mathcal{F}[f](\omega) = \int_{-\infty}^{\infty} f(t)e^{-i\omega t}\, dt$$

The Inverse Fourier Transform (IFT) on the other hand is given by

$$f(t) = \mathcal{F}^{-1}[\hat{f}](t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{f}(\omega)e^{i\omega t}\, d\omega$$

It is also worth recall that the Fourier Transform has also a discrete form, the Discrete Fourier Transform (DFT). In fact, given a sequence $f[n]$ whose series is absolutely convergent, one can define is FT by

$$\hat{f}(\omega) = \mathcal{F}[f](\omega) = \sum_{-\infty}^{\infty} f[n] \cdot e^{i\omega n},$$

and its Inverse Discrete Fourier Transform (IDFT) is given by:

$$f[n] = \mathcal{F}^{-1}[f](t) = \frac{1}{2\pi}\int_{T=2\pi} \hat{f}(\omega) \cdot e^{i\omega n}\, d\omega,$$

where one must remember that the frequency function is over the reals (the time is discrete, not the frequency).

The important aspects to recall here is that the FT gives frequency information about a signal, representing its frequency components and their respective magnitude. Therefore, it is usual to say that the FT has perfect **frequency resolution**. However, FT does not provide any time information as to when those frequencies exist and, therefore, the FT has zero **time resolution**. Hence, FT is a perfect tool for stationary signals, *i.e.*, signals that do not change with time.

To illustrate this situation, Figure 4.1 bellow shows two different signals in time that have the same FT.

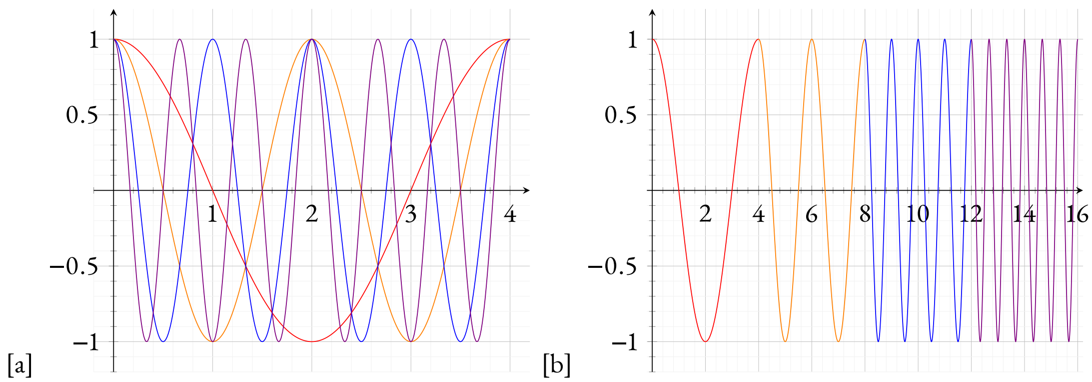

Figure 4.1: Two distinct signals in time that have the same Fourier Transform. In subfigure [a], the signal is given by the superposition of all four signals that appear in the plot, whist in subfigure [b], the signal is given by the concatenation of the four signals.

To remedy this situation for signals that do change with time, sometimes the **Short-Time Fourier Transform (STFT)** is used. The idea of STFT is to imagine that some portion of the non-stationary signal is stationary. The STFT can be defined under the same conditions as the FT, and the central concept is to apply the FT for each stationary portion along the signal and add all of them up.

To be more precise, in the STFT, the original function is multiplied by a **window function** (that is simply the characteristic function of some finite interval†) with some fixed length and the FT is taken with respect to this product. This will be done as the window function moves along the real line, taking the FT at each stationary section. Hence, the STFT will be defined as:

$$\hat{f}(\tau, \omega) = \int_{-\infty}^{\infty} f(t)\mathrm{w}(t-\tau)e^{-i\omega t}\, dt,$$

† There are other possibilities to consider for the window function, but this is usually the easiest to define.

where w denotes the window function and τ is the translational parameter. The conditions for the FT are sufficient to define the STFT because the STFT is essentially the FT defined to the product of the function by the window function. However, due to the parameter τ, the STFT provides some time localization.

Hence, the STFT can be visualized in a three dimensional graphic, where one axis denotes the magnitude, and the other two denote time and frequency.

It is clear from the definition that the STFT carries both uncertainty in time and in frequency. Therefore, there is some resolution in time at a cost of losing precise resolution at frequency. Consequently, with the STFT it is impossible to know what frequencies exists at what time instance and it is only possible to know what frequency bands exists at time intervals. This is again a manifestation of the uncertainty principle (which will be explained in more detail in the following chapters), where an lower bound is in some sense (which will be made more precisely later) given by

$$\Delta t \Delta f \geq \frac{1}{4\pi}.$$

One limitation of the STFT is that the length of the window function is fixed and, therefore, so are both the time and frequency resolutions for the entire signal. This can be visualized in a grid (something that will be very useful to understand another tools to be presented ahead), and it is illustrated in Figure 4.2 bellow.

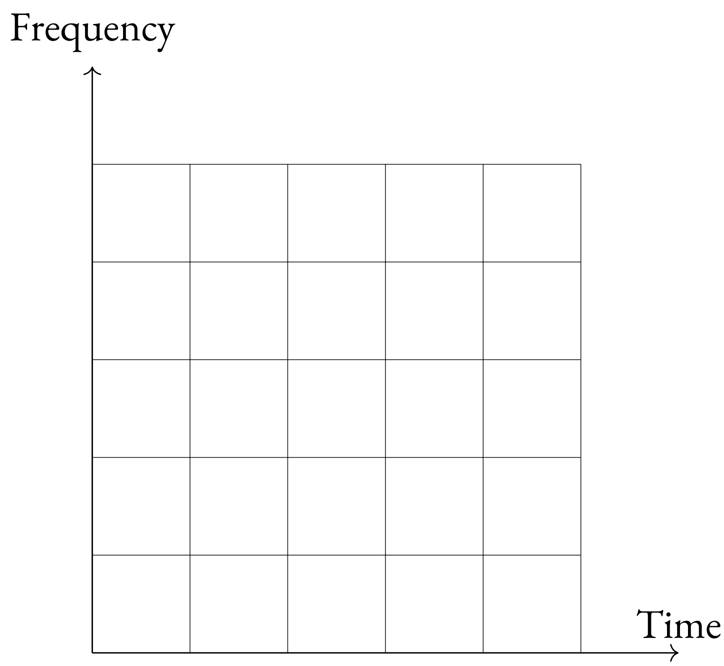

Figure 4.2: Grid in time-frequency domain for the Short Time Fourier Transform.

In the Grid of Figure 4.2 above, a narrow window means good time resolution to the detriment of frequency resolution, while a wide window means bad time resolution to provide good frequency resolution.

The fixed resolution in the time-frequency domain is a often a problem considering that low frequency components of a signal sometimes last long periods of time, requiring a high frequency resolution. On the other hand, high frequency components sometimes appear as short outbreaks, requiring higher time resolution.

The techniques in *Multiresolution Analysis* and *Wavelet Transforms* comes in the urge to remedy this situation. More precisely, the fundamental idea is to analyze a signal into its different frequencies at many distinct resolutions. Hence, the time-frequency domain must be partitioned in different resolutions, for example as illustrated in Figure 4.3 bellow.

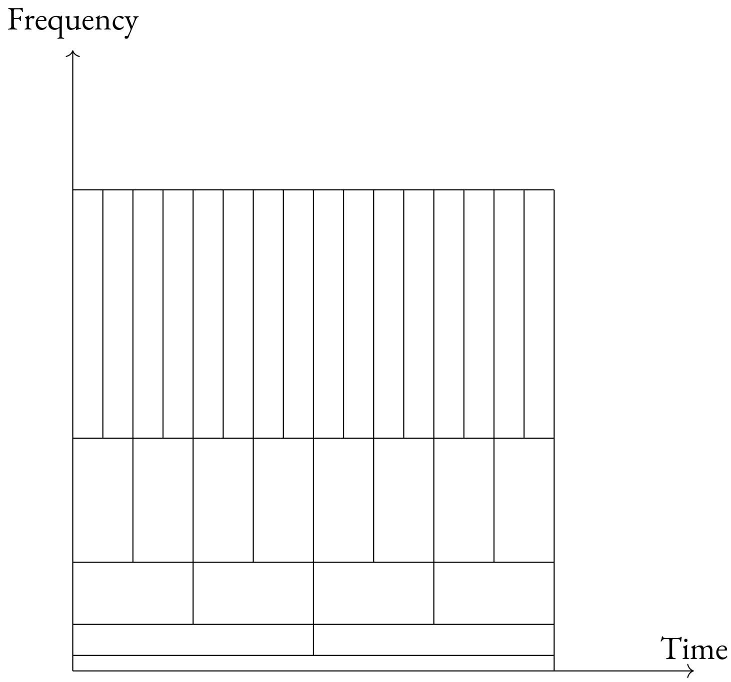

Figure 4.3: Grid in time-frequency domain for the idea of a multiresolution analysis.

The idea shown in Figure 4.3 illustrates a way to solve some of the problems previously discussed, *i.e.*, at high frequencies, there is a good time resolution (at a cost of bad frequency resolution). Likewise, at low frequencies, there is good frequency resolution (at a cost of bad time resolution).

One important insight here is that classical Fourier analysis deals with waves that are infinite. In fact, the complex exponential is a sum of a cosine and a sine. Both these waves (or this one wave if one wants to consider sine as a translation of cosine) have unlimited support in the time domain and they provide, via Fourier Transform, exact frequency resolution. The idea then is to search of a trade off relationship between time resolution and frequency resolution and one idea to start that process is to consider a small wave, the so called *wavelets*[†].

Definition 38. *A* ***wavelet*** *is a piece-wise continuous function that has arbitrarily small spectral power*

[†]The term wavelet is an English word created to adapt the original french word “ondelete”, which means small wave.

outside a sufficiently large interval and its not the zero function (for the discrete case, this will also be the case taking the discrete topology).‡

Therefore, wavelets can be viewed as a "brief oscillation". Some examples are depicted in Figure 4.4 bellow.

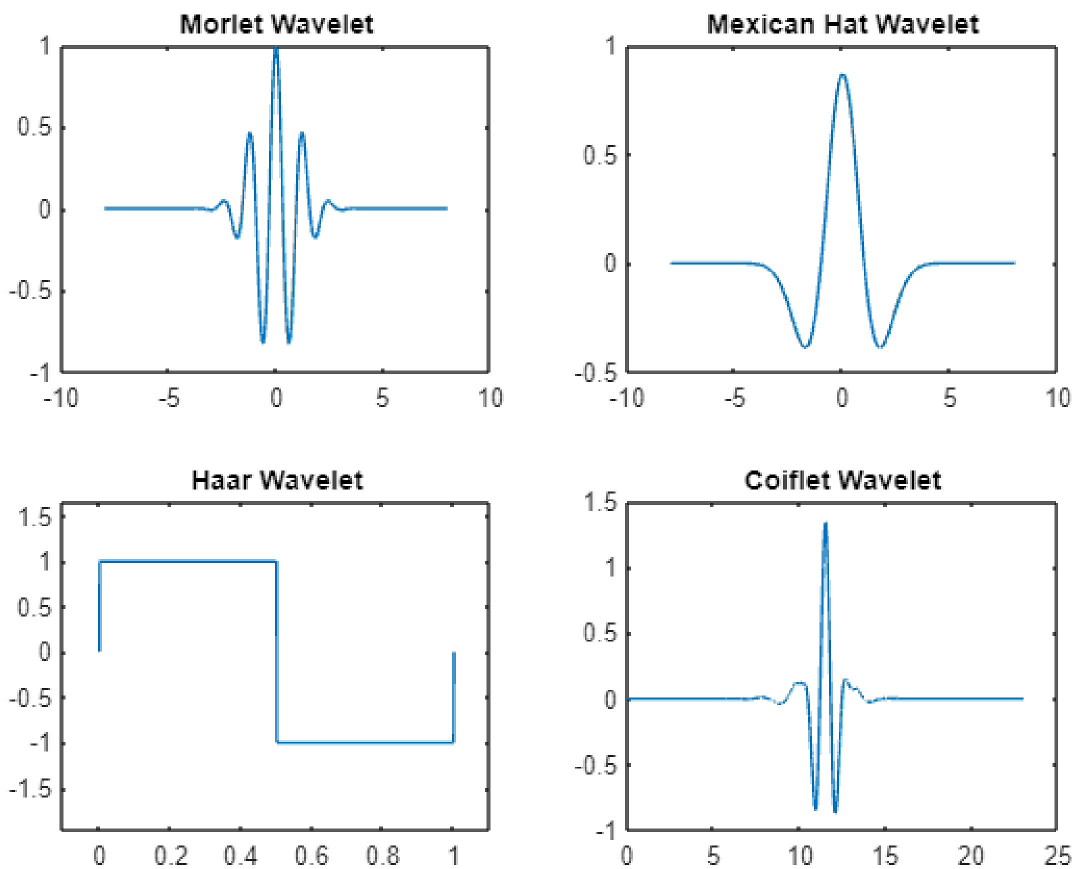

Figure 4.4: Some types of real wavelets with their respective names, generated on MATLAB. They were generated with the parameters wavefun('morl',8); wavefun('mexh',10); wavefun('db1',10); wavefun('coif4',10); respectively

Some of the Wavelets depicted in Figure 4.4 above are actually complex functions, as the exemplified

‡Some authors define wavelets as compactly supported functions. However, some of the most used types of this functions in signal processing, data processing and telecommunications are not compactly supported.

next.

Example 27. *The **Complex Morlet Wavelet** is a function $\Psi : \mathbb{R} \longrightarrow \mathbb{C}$ that can be defined as:*

$$\Psi(t) = \frac{1}{\sqrt{\pi \cdot f_b}} \cdot \exp\big(2\pi i \cdot f_c \cdot t\big) \cdot \exp\left(\frac{-t^2}{f_b}\right),$$

where f_b is the time-decay parameter and f_c the center frequency. This function can be visualized in Figure 4.5 bellow.

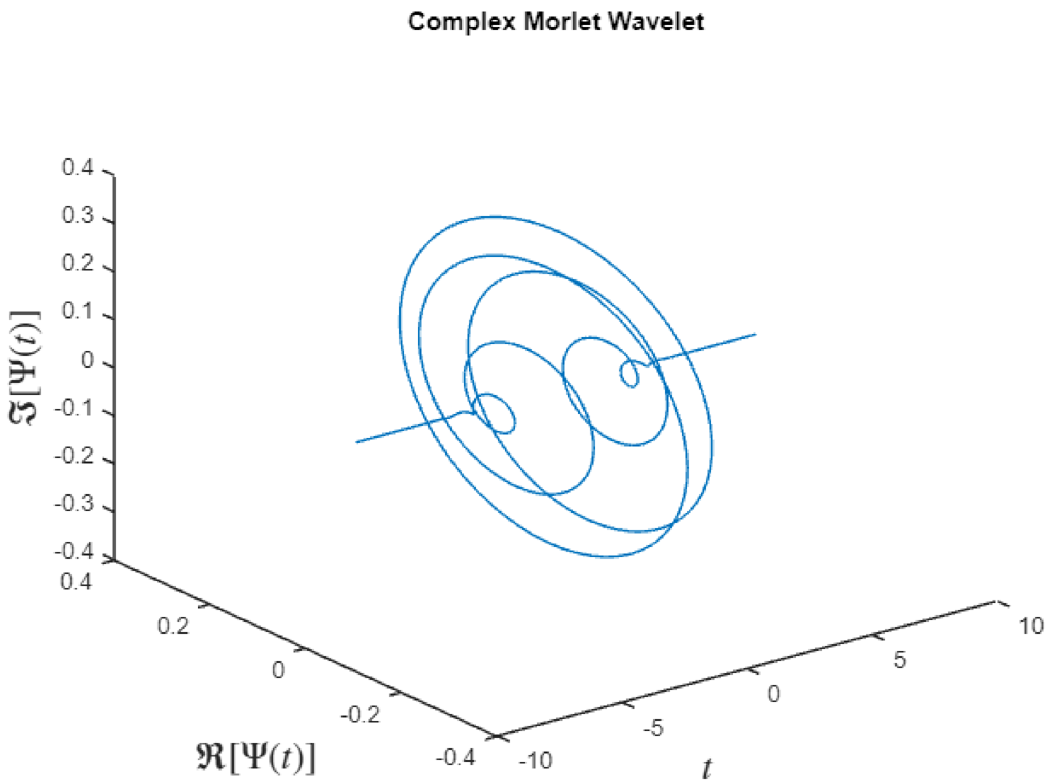

Figure 4.5: Complex Morlet Wavelet plotted in MATLAB. The command is [psi, x]=cmorwavf(Lb, Ub, N, fb, fc); with parameters N=1000; Lb=-8; Ub=8; fb=3; fc=1;, where “fb” is the time-decay parameter, “fc” is the center frequency,

General Morlet wavelets can be defined by Gaussian probability distributions, and are composed by a complex exponential, known as the carrier, and a Gaussian Window, known as the envelope.

Definition 39. *Given a particular wavelet $\psi : \mathbb{R} \longrightarrow \mathbb{C}$, it is possible to define its **Family of Wavelets** that are the set of wavelets defined as translations and compressions of the original wavelet. The original wavelet $\psi : \mathbb{R} \longrightarrow \mathbb{C}$ is known as the **Mother Wavelet**. The other wavelets are defined as*

$$\psi_{a,b} = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right),$$

for $a > 0$ and $b \in \mathbb{R}$.

The **(Continuous) Wavelet Transform** (WT) can then be defined, given a mother wavelet $\psi : \mathbb{R} \longrightarrow \mathbb{C}$ and a function $f : \mathbb{R} \longrightarrow \mathbb{C}$, for all the set of the family as:

$$W_\psi[f](a,b) = \langle f, \psi_{a,b} \rangle = \int_{\mathbb{R}} f(t) \cdot \psi^*_{a,b}(t)\, dt,$$

where the symbol $(*)$ denotes the complex conjugate of the function, *i.e.*, $\psi^*_{a,b}(t) = \overline{\psi_{a,b}(t)}$. Hence, the wavelet transform is given by the standard inner product in $L^2(\mathbb{R})$ of the function f against the members $\psi_{a,b}$ of the family of the mother wavelet ψ.

The idea in the wavelet transform is that the wavelets are the new basis functions instead of complex exponentials (or sine and cosine). However, the wavelets in the family can be translated and scaled. This is important because wide wavelets are better to resolve low frequency components of the signal with bad time resolution, whilst shrunken wavelets are better to resolve high frequency components of the signal with good resolution.

The Inverse Wavelet Transform (IWT) and the Discrete Wavelet Transform (DWT) are a bit more complicated and will be treated in the next sections.

To illustrate the use of the WT in signal processing, the next section will be dedicated to give an example of analysis with some of the tools mentioned in this section before the formal treatment is given in the last sections.

4.2 Applications to Signal Processing

In this section, some of the tools presented in the former section, namely the STFT and the WT, will be used to analyse a particular signal with some specific characteristics and illustrate the advantages and disadvantages of each one.

For that matter, consider a signal (with some very quirk features) displayed in Figure 4.6 bellow.

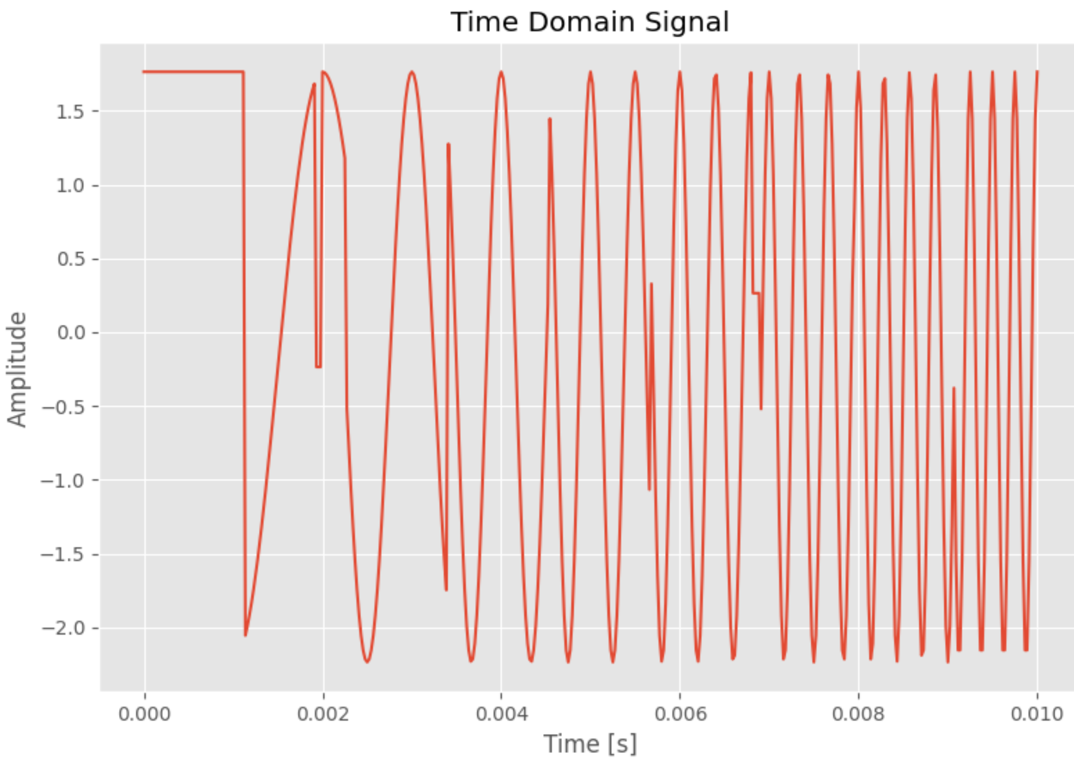

Figure 4.6: Chirp signal with high frequency interjections.

This particular signal would be very difficult to analyse by standard methods as the usual FT. His frequencies varies along time in a very perceptible way and it has sharp edges, which gives more high frequency components to the signal. Hence, in order to obtain the maximum amount of information of the signal, the information with the analysis via STFT and WT will be compared.

As an important observation, since this is done in a computer program, everything will be in essence discrete: the signal, the integrals, the window functions, etc. However, some interpolation is done in the graphics to simulate the continuous case of the STFT and WT. Hence, the size of the window function will be given in term of samples.

Firstly, it is important to compare two diferent STFT's that differ based on the size of the window function. This can be viewed in Figure 4.7, for a window function size 8, and Figure 4.8 for a window function size 256.

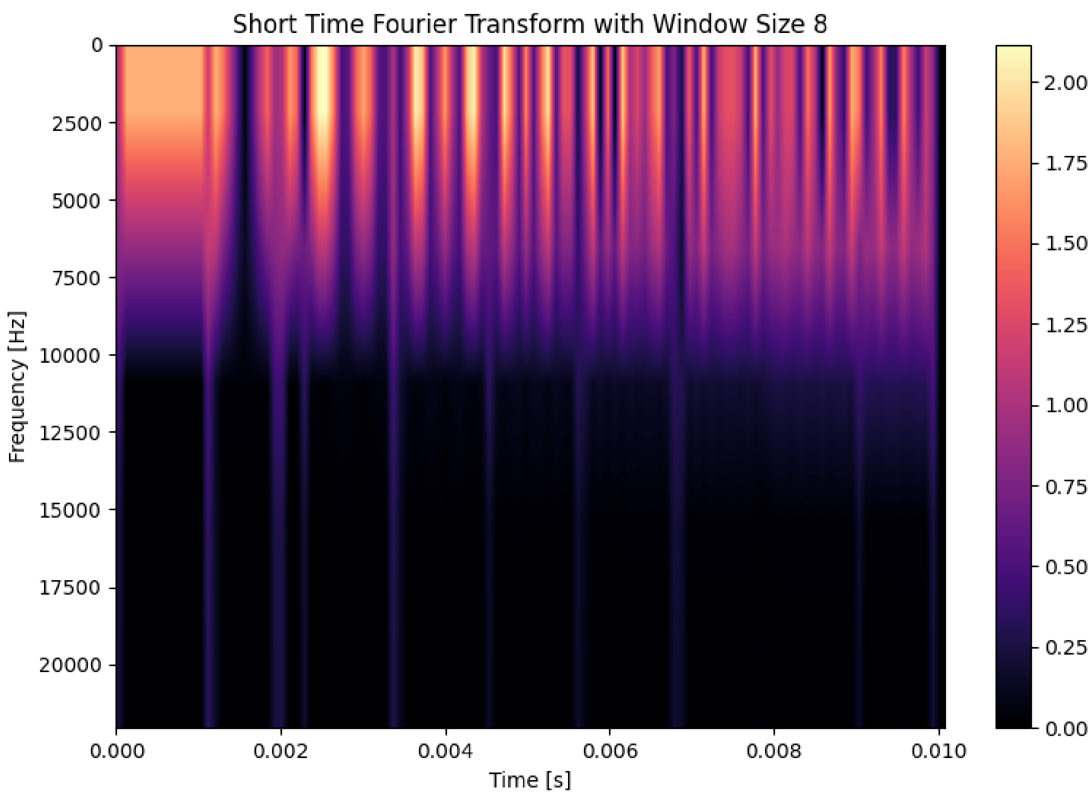

Figure 4.7: Short Time Fourier Transform of the signal with window size 8.

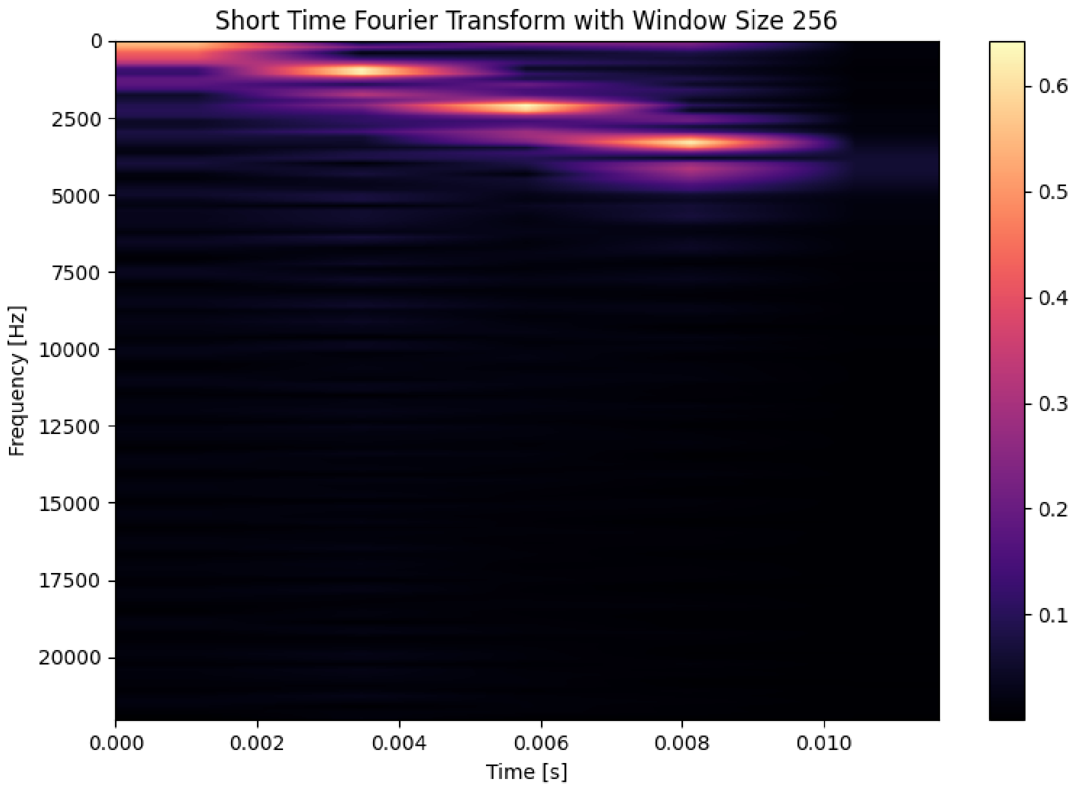

Figure 4.8: Short Time Fourier Transform of the signal with window size 256.

The difference in this images is very clear. Note how in Figure 4.7 the time resolution is much smaller, but at a cost of having a much worse frequency resolution compared with Figure 4.8. Note how individually this single figures do not provide all the important information about the signal. For example, the high frequency components at sharp edges are almost to detect in time impossible to see in Figure 4.8, whist in Figure 4.7 there is a time information about when these high frequency bursts occur, despite having low information about what are really the frequency components involved.

To remedy this situation, the WT was presented in the previous section. Figure 4.9 ahead shows the WT for the signal that its been analyzed.

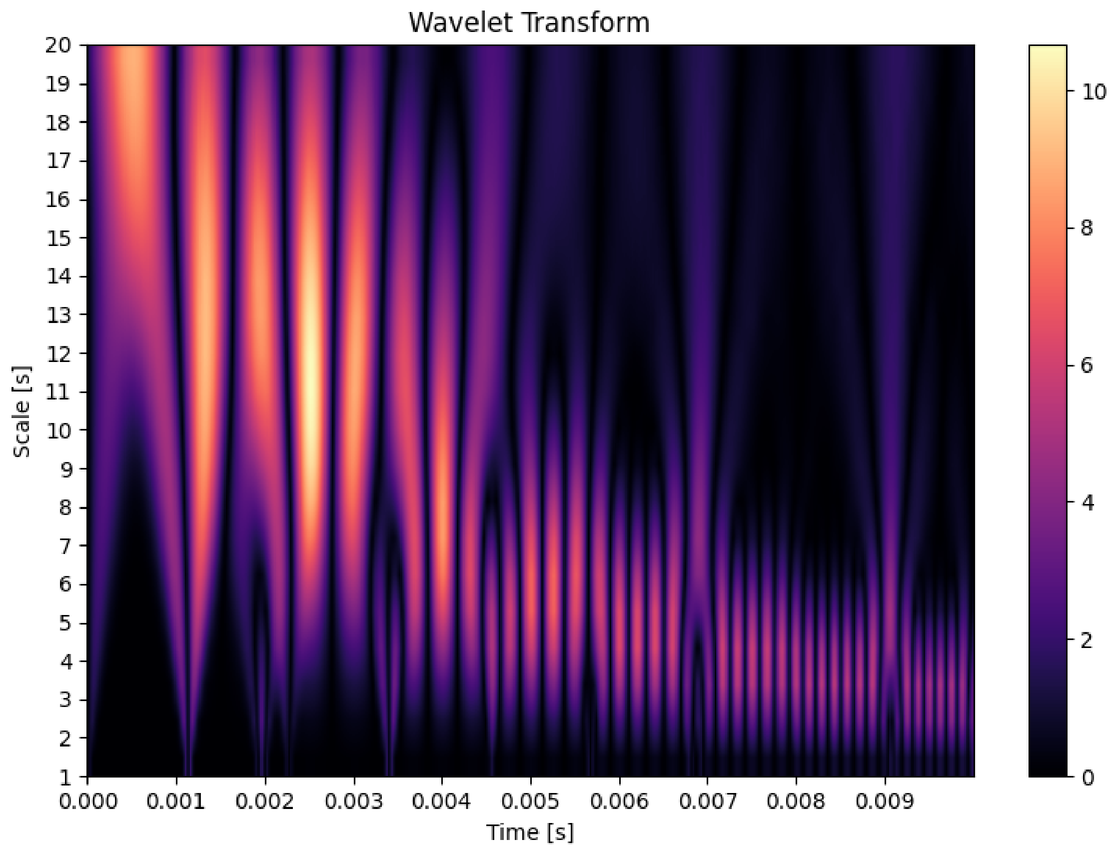

Figure 4.9: Wavelet Transform of the signal

Recall that the WT of a signal f with respect to a mother wavelet ψ is given by:

$$W_{\psi}[f](a,b) = \int_{\mathbb{R}} f(t) \cdot \psi^{*}_{a,b}(t)\, dt,$$

where

$$\psi_{a,b} = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right).$$

Hence, a is the scaling parameter and the time-frequency domain illustrated in Figure 4.3 is varying vertically with the parameter a. Also, it is important to emphasize that the goal of the wavelet transform is to have high time resolution for high frequencies and high frequency resolution for low frequencies. Figure 4.9 shows for low values of frequencies with a high time resolution at the lower part. It is important to note that the vertical axis relates the inverse of the frequency, hence given in seconds.

After seeing the information given in Figures 4.7, 4.8 and 4.9, it is straightforward that the WT provides much more information than the classical STFT for theses types of signals with sharp edges, frequency bursts and frequency variations.

The codes in Python Language to generate these figures and the signal transforms can be found in Appendix B.

4.3 Background For More Rigorous Formulation Of The Wavelet Transform

Many of the concepts here are just a review of the concepts developed in the context of Chapter 1.

Recall that a Banach Space is a normed vector space X over a field $\mathbb{K}$ with norm $||\cdot||$ such that X is complete with respect to this norm (every Cauchy sequence in X converges in X), and the convergence is taken with respect to the distance d_X induced by this norm such that $d_X(x_1, x_2) = ||x_1 - x_2||$. With these tools, it is possible to define infinite linear combinations (infinite sums) by limits:

$$\sum_{n==1}^{\infty} x_n = \lim_{n \longrightarrow \infty} \sum_{k=1}^{n} x_k.$$

A Schauder Basis for a Banach Space X will be a countable sequence $(e_n)_{n=1}^{\infty}$ of elements of X such that every element $x \in X$ can be uniquely expressed as a infinite linear combination

$$x = \sum_{n=1}^{\infty} a_n \cdot e_n,$$

for appropriate coefficients $a_n \in \mathbb{K}$.

Recall that, given a non empty vector space V over a field $\mathbb{K}_V$, then a set β of elements of V is a Hamel basis in V if β is linearly independent and given any element $v \in V$, then there exists a finite subset F such that is possible to write in a uniquely way:

$$\sum_{k \in F} a_k x_k,$$

where $a_k \in \mathbb{K}_V$ and $x_k \in \beta$ for every $k \in F$. The existence of Hamel basis is ensured by Zorn's Lemma. Moreover, given any linearly independent set α, there exists a Hamel basis β such that $\alpha \subseteq \beta$.

Schauder basis differ from Hamel basis and they do not coincide in general. If fact, it is only when one has good enough topology that is possible to talk about Schauder Basis. By these definitions the span with respect to a Schauder Basis and a Hamel basis can be different. This is because Schauder independence is stronger than Hamel independence.

Hence, every space has a Hamel basis, but in fact, not every space has a Schauder basis. Moreover, the basis in the case of Schauder must be ordered, since unconditional convergence is not guaranteed in general.

Example 28. *The sequence spaces ℓ^p with $1 \leq p < \infty$ and c_0 have as canonical basis $(e_n)_{n=1}^{\infty}$ with the elements given by:*

$$e_n = (\delta_{nk})_{k=1}^{\infty}.$$

In fact, every Banach Space with a Schauder basis is separable, but the converse is not true. Even a smaller class of Banach spaces has unconditional basis.

The situation turns out to be a lot more well behaved when the space is a inner product space with some "additional compatibility". Indeed, every vector space over an infinite ordered field (because, for example, finite fields are not ordered and therefore would violate the positive-definiteness of the inner product) admits an inner product. The restriction to work in Hilbert Spaces is a good way to remedy this situation.

Recall that a Hilbert Space is an inner product space $(H, \langle \cdot, \cdot \rangle)$ which is complete with respect to the norm $\|x\| = \sqrt{\langle x, x \rangle}$ (and therefore a Banach space with respect to this norm). It is important to remember that every Hilbert Space H has an orthonormal basis in a sense involving convergence (see Chapter 1), which will be a maximal (with respect to the canonical subset ordering) orthogonal subset $\beta \subseteq H$ (a Hilbert basis). The most important examples here will be the spaces $L^2([0,1])$ and $L^2(\mathbb{R}^n)$ with the canonical inner product:

$$\langle f, g \rangle = \int_H f \cdot \overline{g}\, d\lambda,$$

where λ is the Lebesgue measure.

Then, given a Hilbert basis β for H, every element $x \in H$ can be written as:

$$x = \sum_{b \in \beta} \langle x, b \rangle b,$$

where is understood that there are at most a countable collection of nonzero terms and for any enumeration $\{b \in \beta : \langle x, b \rangle \neq 0\} = \{b_1, \ldots\}$ of nonzero elements:

$$x = \lim_{n \longrightarrow \infty} \sum_{j=1}^{n} \langle x, b_j \rangle b_j.$$

Hence, the norm can be represented as:

$$\|x\| = \sqrt{\sum_{b \in \beta} |\langle x, b \rangle|^2}$$

Consider now $\mathbb{C}$-vector spaces. The space $L^2([0,1])$ has then a canonical basis $\{e_n\}_{n \in \mathbb{Z}}$ given by

$$e_n(t) = e^{2\pi i n t}.$$

Hence, every $f \in H = L^2([0,1])$ can be written as

$$f = \sum_{n \in \mathbb{Z}} \langle f, e_n \rangle e_n,$$

where the convergence with respect to the L^2-norm. This is the Fourier series constructed in Chapter 1. It is usual to write

$$\hat{f}(n) = \langle f, e_n \rangle = \int_0^1 f(t) \cdot e^{-2\pi int}\, dt.$$

The partial symmetrical sums will be denoted by

$$S_n f(t) = \sum_{k=-n}^{n} \hat{f}(k) e^{2\pi ikt}.$$

Hence, these partial sums approximate any function $f \in L^2([0,1])$:

$$\lim_{n \longrightarrow \infty} \left\| f - S_n f \right\|_2 = 0.$$

However, it is worth remember that almost everywhere convergence (a.e.-convergence) implies L^p-convergence, and in fact, Calerson-Hunt's Theorem states that

$$S_n f(t) = \sum_{k=-n}^{n} \hat{f}(k) e^{2\pi ikt} \xrightarrow[n \longrightarrow \infty]{a.e.} f(t),$$

for every periodic function in L^p, $1 < p < \infty$ (in this context, $L^p([0,1])$). Hence, the exponential form a Schauder basis for $L^p([0,1])$, $1 < p < \infty$.

In the case of $p = 1$, the exponentials e_n do not form a basis for $L^1([0,1]) \longrightarrow \mathbb{K}$ or for the space $C([0,1])$ of continuous functions on the interval with the supremum norm. Nonetheless, given $f \in L^1([0,1])$, there is enough information in the Fourier coefficients $(\hat{f}(n))_{n \in \mathbb{Z}}$ to reconstruct the function f. The real difficulty is know if a given sequence $(b_n)_{n\in\mathbb{Z}}$ is the sequence of some unknown function $g \in L^1([0,1])$.

The basis composed by the exponential functions is indeed canonical and very good basis to deal with (see Chapter 1). One common way to try to justify this choice is that the complex exponentials are "eigenvectors" for the differential operator $\frac{d}{dt}$. However, this may lead to a problem both because this operator is not defined in all $L^2([0,1])$ and because there are other "eigenvectors", for example $e_z(t) = e^{2\pi izt}$ for $z \in \mathbb{C} \setminus \mathbb{R}$.

One way to try to bypass this situation is to argue that actually that the right context to work with this tools is to deal with periodic functions with unitary period (1-periodic). Hence, given a function $\tilde{f} \in L^2([0,1])$ (that is defined only in $[0,1]$), it is natural to extend $\tilde{f}$ to a function f defined on all $\mathbb{R}$ by discarding the value at 1 or 0 and putting $f(t) = \tilde{f}(t - \lfloor t \rfloor)$. Hence, the natural basis elements for the Fourier series, e_n, are 1-periodic and "eigenvalues" of the differentiation.

Another way to bypass this is to argue that actually one is dealing with functions defined on the unit circle $S^1 = \mathbb{T}$ since $\mathbb{T}$ can be expressed as $\mathbb{T} = \{e^{2\pi it} : t \in \mathbb{R}\} = \{e^{2\pi it} : t \in [0,1)\}$ and then the e_n would be viewed as unitary irreducible representations of $\mathbb{T}$. However, this is a very particular case of the Fourier Analysis on topological groups done in Chapter 1, where was considered the case for $L^2(G)$ and G is a locally compact abelian group. The space L^2 is defined with respect to the Haar measure on G (and normalized to give a probability in the case where G is compact). Then, the general Fourier Series of $f \in L^2(G)$ will be indexed by the set of characters in $\hat{G}$.

Throughout the next discussion, where the goal is to recall the definition of Fourier Transforms for this context, it is actually useful to consider, even though the context is of $\mathbb{C}$-valued vector spaces, that a function $f \in L^2$ that is gonna be represented is real valued. This is just to simplify the calculations. The results in the parts where this is done can be reformulated considering the real and imaginary part and the result follows from linearity.

Next, one may consider Fourier Series defined on general intervals $[a,b]$, $a < b \in \mathbb{R}$. This can be simply done by a translational operator that does a change of variables:

$$\begin{aligned} T_{a,b} : L^2([0,1]) &\longrightarrow L^2([a,b]) \\ f &\longmapsto f_{a,b} : [0,1] \longrightarrow [a,b] \\ & \qquad s \longmapsto \frac{1}{\sqrt{b-a}} f\left(\frac{s-a}{b-a}\right). \end{aligned}$$

Indeed, $T_{a,b}$ is an isometry and hence preserves inner product, which enables to transfer the Fourier Series from one interval to another.

Now, if $[a,b] = [-T,T]$, $T \in \mathbb{R}_+^*$, then given any $f \in L^2([-T,T])$,

$$f(t) = \sum_{n\in\mathbb{Z}} \hat{f}(n) \left(\frac{(-1)^n}{\sqrt{2T}} \exp\left(\frac{2\pi i n t}{2T}\right) \right),$$

where $\hat{f}(n)$ is given by

$$\hat{f}(n) = \frac{(-1)^n}{\sqrt{2T}} \int_{-T}^{T} f(t) \exp\left(\frac{-2\pi i n t}{2T}\right) dt$$

Hence, these last two relations lead to the Fourier Series explicitly in terms of f:

$$f(t) = \sum_{n\in\mathbb{Z}} \frac{1}{2T} \left(\int_{-T}^{T} f(x) \exp\left(\frac{-2\pi i n x}{2T}\right) dx \right) \exp\left(\frac{2\pi i n t}{2T}\right),$$

where the variable x does not change anything and was introduced just to avoid repetition on t. Hence, if f is a function with compact support in $L^2(\mathbb{R})$, there is a sufficiently large value of T such that:

$$f(t) = \sum_{n\in\mathbb{Z}} \frac{1}{2T} \left(\int_{-\infty}^{\infty} f(x) \exp\left(\frac{-2\pi i n x}{2T}\right) dx \right) \exp\left(\frac{2\pi i n t}{2T}\right).$$

Now define the function $\hat{f} = \mathcal{F}[f]$ by:

$$\hat{f}(\omega) = \mathcal{F}[f](\omega) = \int_{-\infty}^{\infty} f(t)e^{-2\pi i\omega t}\,dt.$$

Note that there is a factor 2π in the exponential, which is different from what was presented at Section 4.1, but this does not affect the general results and is just a convention to consider, for example, characters in the unity circle $\mathbb{T}$. Hence:

$$f(t) = \sum_{n\in\mathbb{Z}} \frac{1}{2T}\mathcal{F}[f]\left(\frac{n}{2T}\right)\exp\left(\frac{2\pi int}{2T}\right) = \sum_{n\in\mathbb{Z}} \frac{1}{2T}\hat{f}\left(\frac{n}{2T}\right)\exp\left(\frac{2\pi int}{2T}\right).$$

Considering this as a Riemann (or Stieltjes if one wants a bit more generality) sum approximating the integral and taking the limit for $T \longrightarrow \infty$:

$$f(t) = \int_{-\infty}^{\infty} \hat{f}(\omega)e^{2\pi i\omega t}\,d\omega,$$

for all $f \in L^2(\mathbb{R})$ compactly supported. But, as seen in Chapter 3, this is an isometry in $L^2(\mathbb{R})$, so that this formula actually works for all $f \in L^2(\mathbb{R})$. Indeed, recall that due to Parseval's theorem, this is in fact an isometric isomorphism and admits and adjoint $\mathcal{F}^{-1}$.

In the previous case of Fourier Series, the context was of functions defined on $[0, T)$, T-periodic functions or elements of the circle of radius 1. There, the functions were expressed as super-positions of a countable collection of exponentials of the form $e^{2\pi int/T}$, which have multiples of T as periods. For functions defined on all $\mathbb{R}$, the number of periods will be uncountable, given by all defined values of $1/\omega$. Remember that in the case of G Hausdorff compact, the dual group $\hat{G}$ will be countable if and only if G is second countable (and, since G is Hausdorff, will be metrizable as well). Hence, the exponentials occur as a superposition to define every $f \in L^2(\mathbb{R})$, as before, but an uncountable one. These exponentials will also be the all the bounded eigenfunctions for the derivative operator on $\mathbb{R}$, which is a reason Fourier Transforms are so used in the theory of differential equations.

Note furthermore that the exponential functions $t \longmapsto e^{2\pi i\omega t}$ are periodic functions with period $1/|\omega|$. Hence, it is usual to say that this function "repeats it self" $|\omega|$ times as t increases to $t+1$, which is one of the main reasons to call $|\omega|$, or even ω the frequency, and hence, call the codomain the Frequency Domain.

4.4 Shannon-Nyquist Sampling Theorem, Limited Bands and Heisenberg's Uncertainty Principle

In the context of telecommunications, the concept of *Band Limited* signals, *i.e.*, signals that have only a finite interval on frequency domain for which the signal is nonzero, is very important for numerous reasons. One of them is that most circuits work as filters and they do not allow every frequency to bypass or interfere in the system, which is good most of the times, because noise usually comes in the form of high frequency perturbations, so having a smaller pass-band is positive to avoid too much noise. Another reason is that the number of channels available for communications is restricted, so many users must share the same channel and the way they do that is often to choose different frequencies for the **carrier**, which is a signal used to code, transmit and decode the original signal. The process of passing the information of the signal to the carrier can be done in numerous ways, for instance frequency modulation (FM), amplitude modulation (AM) and phase modulation (PM).

For this to be possible, *i.e.*, many users share the same channel, the band limitation of the signal is essential to avoid interference and to be possible to later on decode the signal.

Now consider a band-limited signal. Without loss of generality, it is possible to suppose that this signal is limited in the interval $[-B_w/2, B_w/2]$, where B_w stands for Band width. If this is not the case, a simple multiplication by a complex exponential $e^{2\pi i\omega_0 t}$ with a suitable ω_0 will do the job. This is easy to remember if one remember that the Fourier transform of a exponential in time is the Dirac's Delta and that multiplication in time corresponds to convolution in frequency domain. Hence, doing a convolution by the $\delta(\omega_0)$, by the sampling property, corresponds to translate the signal to be centered at ω_0.

Now, consider a signal $f \in L^2(\mathbb{R})$ that is compactly supported. If $g = \hat{f} = \mathcal{F}[f]$, then g can be reconstructed by the Fourier series (note that here the assumption of the Fourier transform being real is used, but can be extended by linearity to the complex case):

$$g(\omega) = \sum_{n\in\mathbb{Z}} \hat{g}(n) \left(\frac{(-1)^n}{\sqrt{B_w}} \exp\left(\frac{2\pi i n\omega}{B_w}\right)\right) \qquad (|\omega| < B_w)$$

where $\hat{g}(n)$ is given by

$$\begin{aligned}
\hat{g}(n) &= \frac{(-1)^n}{\sqrt{B_w}} \int_{-B_w/2}^{B_w/2} g(x) \exp\left(\frac{-2\pi i n x}{B_w}\right) dx \\
&= \frac{(-1)^n}{\sqrt{B_w}} \int_{-B_w/2}^{B_w/2} \mathcal{F}[f](x) \exp\left(\frac{-2\pi i n x}{B_w}\right) dx \\
&\overset{\star}{=} \frac{(-1)^n}{\sqrt{B_w}} \int_{-\infty}^{\infty} \mathcal{F}[f](x) \exp\left(\frac{-2\pi i n x}{B_w}\right) dx,
\end{aligned}$$

where the equality in $(\star)$ uses the fact that the signal is band-limited. But since $g = \hat{f} = \mathcal{F}[f]$, one may think that this should be $(1/\sqrt{B_w})f(-n/B_w)$ because of the inversion formula. However, this is the same point discussed to the exhaustion at Chapter 1. Indeed, the Fourier Inversion Formula works in $L^2(\mathbb{R})$ and thus, can not be applied pointwise. However, one should recall the discussion, made at the section of Fourier on groups, about the existence of a continuous representative for the equivalence class at L^2. Here, it turns out that in fact the band limited functions are necessarily continuous, and, therefore, the inversion formula works pointwise. Thus, in this case, $\hat{g}(n)$ will indeed be $(1/\sqrt{B_w})f(-n/B_w)$. Hence, using the Fourier Series formula (the factors $(-1)^n$ cancel each other out):

$$\hat{f}(\omega) = \mathcal{F}[f](\omega) \sum_{n\in\mathbb{Z}} f\left(\frac{-n}{B_w}\right)\left(\frac{1}{B_w}\exp\left(\frac{2\pi i n\omega}{B_w}\right)\right) \qquad (|\omega| < B_w).$$

With this results and combining the Fourier Inversion formula and the fact that the signal is band-limited, it is possible to prove the following result:

Theorem 23. *Let $f \in L^2(\mathbb{R})$ be a band-limited signal with $\mathcal{F}[f]$ supported in $[-B_w/2, B_w/2]$. Then f is completely determined by its values $f(n/B_w)$, $n \in \mathbb{Z}$ and moreover:*

$$f(t) = \sum_{-\infty}^{\infty} f\left(\frac{n}{B_w}\right)\frac{\sin(\pi(n - B_w t))}{\pi(n - B_w t)} = \sum_{-\infty}^{\infty} f\left(\frac{n}{B_w}\right) sinc(\pi(n - B_w t)).$$

Hence, if the range of frequencies of the signal f is B_w, then one can completely reconstruct the signal by sampling with frequency $2B_w$.

Proof. Without loss of generality, suppose that $B_w = 1$. Otherwise, a simple scaling or change of units will do the job. Hence, the signal f will have band-width supported in $[-1/2, 1/2]$, *i.e.*, it will be nonzero only for $|\omega| < 1/2$.

Let g be a periodic function with period 1 such that g coincides with $\hat{f}$ for $|\omega| < 1/2$. The idea is to prove that the coefficients of g coincide with the values $f(n)$ for $n \in \mathbb{Z}$. If the values of f ate the integers are known then it is possible to know the g which gave $\hat{f}$. Therefore, inverting the Fourier Transform it is possible to recover f.

Begin expanding g in its Fourier series:

$$g(\omega) = \sum_{-\infty}^{\infty} c_n e^{-2\pi i n\omega}, \qquad (\clubsuit)$$

where the coefficients are given by:

$$c_{-n} = \underbrace{\int_{-1/2}^{1/2} g(\omega)e^{-2\pi i n\omega}\,d\omega}_{[A]} = \underbrace{\int_{-1/2}^{1/2} \hat{f}(\omega)e^{-2\pi i n\omega}\,d\omega}_{[B]} = \underbrace{\int_{-\infty}^{\infty} \hat{f}(\omega)e^{-2\pi i n\omega}\,d\omega}_{[C]} = f(n). \qquad (\blacklozenge)$$

Recall that g is a periodic function, and thus, have a Fourier series, not the Transform with frequencies in all ($\mathbb{R}$). The Fourier coefficients of g are given in the frequencies that are multiples of the fundamental frequency. The equality [A]=[B] is justified because g coincides with $\hat{f}$ for $|\omega| < 1/2$ and the equality [B]=[C] is justified because $\hat{f}$ is nonzero only in $|\omega| < 1/2$. Finally, by Parseval's Theorem, the Inverse Fourier Transform will give the values of $f(n)$.

The part of the process to uniquely determine f is done. More specifically, if the values of f in $\mathbb{Z}$ are know, then, measuring the signal once at each unity time (which is twice the band-width maximum frequency), it is possible to reconstruct f.

The process to in fact reconstruct f, however, is not yet clear. The goal now is to prove the reconstruction formula stated at the theorem.

Since $\hat{f}$ is band-limited, write f as the Inverse Fourier Transform of its Fourier Transform:

$$f(t) = \int_{-\infty}^{\infty} \hat{f}(\omega)e^{-2\pi i\omega t}\,d\omega.$$

However, $\hat{f}$ is supported in $[-1/2, 1/2]$, which implies that

$$f(t) = \int_{-1/2}^{1/2} \hat{f}(\omega)e^{-2\pi i\omega t}\,d\omega.$$

Since in $[-1/2, 1/2]$, $\hat{f}$ is given by ($\clubsuit$), it is possible to substite this sum in [A] for $\hat{f}$ leading to

$$f(t) = \int_{-1/2}^{1/2} \left(\sum_{n=-\infty}^{\infty} c_{-n} e^{2\pi i n\omega} e^{-2\pi i\omega t} \right) d\omega,$$

where the minus sign at the exponent of (♣) was disregarded since the sum is infinite on the integers and this will not affect the result.

Using (♦), the coefficients c_{-n} can be substituted by $f(n)$ and, by Lebesgue's Dominated Converge Theorem, the order of the integral and the sum can be exchanged:

$$f(t) = \sum_{-\infty}^{\infty} \left(\int_{-1/2}^{1/2} f(n) e^{2\pi i\omega(n-t)} \, d\omega \right) = \sum_{n=-\infty}^{\infty} \left(f(n) \int_{-1/2}^{1/2} e^{2\pi i\omega(n-t)} \, d\omega \right)$$

The integral at the end can be expressed more directly as:

$$\int_{-1/2}^{1/2} e^{2\pi i\omega(n-t)} \, d\omega = \left[\frac{1}{2\pi i(n-t)} e^{2\pi i\omega(n-t)} \right]_{-1/2}^{1/2} = \frac{\sin(\pi(n-t))}{\pi(n-t)} = \operatorname{sinc}(\pi(n-t)),$$

which gives:

$$f(t) = \sum_{-\infty}^{\infty} f(n) \frac{\sin(\pi(n-t))}{\pi(n-t)} = \sum_{-\infty}^{\infty} f(n) \operatorname{sinc}(\pi(n-t)).$$

□

It could be natural, after all this discussion about compactly supported Fourier Transforms to try to consider compact supported functions in the time domain. However, this can not be done because it turns out that it is impossible to have both f and $\mathcal{F}[f]$ compactly supported. This fact follows from the following theorem that relates complex analysis and distribution theory:

Theorem 24 (Schwartz-Paley-Wiener). *An entire function (complex-valued function that is holomorphic on the hole complex plane) F defined on $\mathbb{C}^n$ is the Fourier Transform of a distribution T of compact suport if and only if for all $z \in \mathbb{C}^n$*

$$|F(z)| \leq C(1+|z|)^N e^{B|\operatorname{Im}(z)|},$$

for some constants C, N and B. Moreover, this distribution T will be supported in the closed ball of center

0 *and radius B and, if for every positive integer M there is a constant* C_M *such that for all* $z \in \mathbb{C}^n$

$$|F(z)| \leq C_M(1 + |z|)^{-M} e^{B|\operatorname{Im}\{z\}|},$$

then T is an infinitely differentiable function (actually a distribution induced by a function of such type) and vice versa.

Proof. See [11]. □

As a particular case for this discussion, there is the result bellow:

Corollary 6 (Paley-Wiener)**.** *Let f be a compactly supported function in* $L^2(\mathbb{R})$*. Then its Fourier Transform* $\mathcal{F}[f] : \omega \in \mathbb{R} \longmapsto \mathcal{F}[f](\omega)$ *extends to be an analytic function* $\omega \longmapsto \mathcal{F}[f](\omega) : \mathbb{C} \longrightarrow \mathbb{C}$ *(hence, an entire function). Moreover, this function must be of exponential type, i.e.,*

$$|\mathcal{F}[f](\omega)| \leq Ae^{B|\omega|} \qquad (\omega \in \mathbb{C})$$

for appropriate constants $A, B > 0$*. Furthermore, the restriction to* $\mathbb{R}$ *of entire functions of exponential type are precisely the Fourier Transforms of compactly supported functions in* $L^2(\mathbb{R})$*.*

Proof. It is a straightforward consequence of the Theorem 24. □

There is an analogue version of Corollary 6 for the Inverse Fourier Transform concerning the last part of Theorem 24, *i.e.*, the band-limited functions are the restrictions to $\mathbb{R}$ of entire functions of exponential type.

However, an entire function that is not identically zero can not vanish on any interval of positive length (see [17]), and hence f and $\mathcal{F}[f]$ can not be both compactly supported.

The Theorem 24 and Corollary 6 provided qualitative information about the fact that both the function in time domain and its Fourier Transform in frequency domain can not be both compactly supported at the same time, but they did not provide any quantitative information.

The quantitative information will be provided next and it is known as the *Uncertainty Principle*. However, this quantitative version will concern functions that have some "localization" in time and whose Fourier Trnasforms are also "localized" in a sense that will be made precise.

Let $f \in L^2(\mathbb{R})$ be a normalized function to have unitary norm, *i.e.*, $\|f\|_2 = 1$. Then f can be viewed as a probability density function on $\mathbb{R}$. The mean μ and the variance σ^2 of f will then be given by:

$$\begin{cases} \mu = \displaystyle\int_{\mathbb{R}} t|f(t)|^2\,dt, \\ \sigma^2 = \displaystyle\int_{\mathbb{R}} (t-\mu)^2|f(t)|^2\,dt. \end{cases}$$

This will be possible to calculate for compactly supported functions in $L^2(\mathbb{R})$, but it turns out to be possible for general normalized functions $f \in L^2(\mathbb{R})$ as well.

Since the Fourier Transform is an isometry in $L^2(\mathbb{R})$, $\left\|\hat{f}\right\|_2 = \|\mathcal{F}[f]\|_2 = \|f\|_2 = 1$, and hence it is also possible to calculate the mean $\hat{\mu}$ and the variance $(\hat{\sigma})^2$ for the Fourier Transform.

Theorem 25 (Heisenberg's Uncertainty Principle)**.** *If* $f, f', f'', tf, t^2 f \in L^2(\mathbb{R})$ *with* $\|f\| = 1$*, then*

$$\sigma\hat{\sigma} \geq \frac{1}{4\pi},$$

or, equivalently,

$$\underbrace{\left(\int_{-\infty}^{\infty} (t-\mu)^2|f(t)|^2\,dt\right)}_{\textit{variance of } t}\underbrace{\left(\int_{-\infty}^{\infty} (\omega-\hat{\mu})^2|\hat{f}(\omega)|^2\,d\omega\right)}_{\textit{variance of } \omega} \geq \frac{1}{16\pi^2}$$

Proof. See [1]. □

4.5 Window Functions And The Short Time Fourier Transform

As shown in the introduction of this chapter, Fourier Transforms wreck the localization of the information that they provide, *i.e.*, they provide the magnitude of the frequency components but not when in time they occur. More over, being an isometric isomorphism in L^2, under certain hypothesis, they enable the reconstruction of the functions in L^2. In fact, under certain hypothesis as well, for some values of p, it is also possible to recover the functions, but not with the same inversion formula that makes L^2 such a natural space to work with. In all those cases, the role transform is required to reconstruct the functions.

In fact, it may be the case that a small localized change in function on time domain will affect meaningly the Fourier Transform and thus, all the Transform is required to reconstruct the function. This is show in the following example:

Example 29. *Consider a interval* $[0, a) \subseteq \mathbb{R}$, $a > 0$, *and its characteristic function* $\mathbb{1}_{[0,a)}$. *Then, calculating the Fourier transform:*

$$\mathcal{F}[\mathbb{1}_{[0,a)}](\omega) = \frac{1 - e^{-2\pi i a\omega}}{2\pi i\omega}.$$

It is clear that, small shifts in the value of a *can affect drastically the values of* $\mathcal{F}[\mathbb{1}_{[0,a)}](\omega)$ *for large values of* ω *for example.*

Similar examples prove that small changes in frequency domain can lead to very different time functions when the Inverse Fourier Transform is taken.

As seen in the Introduction of this Chapter (Section 4.1), the STFT comes to try to remedy the problem of time localization. More precisely, if g is a particular non identically zero function of compact support, the window function, such as $g(t) = \mathbb{1}_{[-r,r]}$, $r > 0$, or a bump function, then the STFT consists of making translations of g, $t \longmapsto g(t-a)$, and take the Fourier transform of these dislocated window functions against the function f considered. Then, one can define the STFT, as the usual Fourier Transform of this product, attempting to provide time localization (and this transform will clearly depend on the value of a as well):

$$\mathcal{SF}[f](a,\omega) = \mathcal{F}[f(t)g(t-a)](\omega) = \int_{\mathbb{R}} f(t)g(t-a)e^{-2\pi i\omega t}\, dt,$$

where $\mathcal{SF}$ stands for Short Fourier to abbreviate the Short Time Fourier Transform.

Of course the choice of g in most cases is not naive, and some good properties are chosen, such as smoothness, continuity by parts, etc.

The STFT provides in fact some time localization, but it is not perfect for localization. Indeed, it depends on the choice of another parameter, a, and the shape of the graph of the window function g can affect the magnitude of the frequency components in the transform. Moreover, this localization is limited by the Heisenberg's Uncertainty Principle, *i.e.*, forcing more time accuracy implies in more uncertainty in the frequency domain. The difference to the usual Fourier Transform is clear, because the FT provides the exact frequency components, whilst the STFT provides frequency bands. Forcing a small value of σ implies large values of $\hat{\sigma}$ and the best scenario occurs when window functions are generated to have its width proportional to the reciprocal of $\hat{\sigma}$.

Furthermore, the ability to detect certain frequencies depend on the width of the window function. More precisely, for high frequencies ω such that the corresponding period $1/\omega$, $\omega \neq 0$, is smaller than the width of the window, the STFT is reasonable to detect the presence of the frequency ω in such an interval. However, when the frequency ω is small (and its respective period $1/\omega$ is large), this will not be the case, since the window will not be large enough to detect variations of f at that frequency. Again, this can be viewed as a consequence of Heisenberg's Uncertainty Principle: Choosing a large window (low time accuracy), it is possible to better resolve the frequencies.

4.6 Continuous Wavelet Transform

The Wavelet Transform still arises from the motivation to provide some localization in time and frequency. However, as explained at the end of the last section, for that to be the case, *i.e.*, for one to obtain good information about some frequency component, the width of the window function can not be totally arbitrary. In fact, to know some localization in time when a certain frequency ω occurs in the function f, the length of the window function must be somewhat close to the associate period $1/\omega$.

The idea naturally is to try to use different windows to analyse the function f at different frequencies. However, due to Heisenberg's Uncertainty Principle, this will also imply a lower time resolution when the window, that in this case will be a wavelet, is chosen to have a wider range.

At the introduction, the definition of wavelets was given. However, like in the case of the STFT, the choice of the wavelet is important and should be smart to make the information about the function that is being analysed more easy to decode.

Many analysis can be done but, in this section, the wavelets ψ will be compactly supported functions in $L^2(\mathbb{R})$. Since there exists a finite closed interval I (hence with finite Lebesgue measure) such that $\operatorname{supp}(\psi) \subseteq I$, then $L^2(\mathbb{R}) \subseteq L^1(\mathbb{R})$. This comes from a more general fact that, given a measure space $(X, \mathcal{X}, \mu)$ with $\mu(X) < \infty$, then, for $1 \leq q < p < \infty$, one has $L^p(X, \mathcal{X}, \mu) \subseteq L^q(X < \mathcal{X}, \mu)$. Hence, $\psi \in L^1(\mathbb{R})$, and here, another requirement will be imposed: $\int_{\mathbb{R}} \psi(t)\, dt = 0$, *i.e.*, ψ will have mean zero. Other properties such as smoothness and continuity by parts will be often desired as well and will depend on the context.

The idea of using different window lengths translates in the fact that, other than in the STFT case, in the Continuous Wavelet Transform (WT), the window functions (wavelets) will not only be translated but will also be scaled in the hope to obtain more information for different values of frequencies ω.

Given a particular function ψ that will be the mother wavelet, as discussed in the introduction, its family of wavelets will be stretched and translated versions:

$$\psi_{a,b} = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right) \qquad (a \in \mathbb{R}_+^*, b \in \mathbb{R}).$$

The argument of ψ in the definition of $\psi_{a,b}$ is a translation by b to the right and a compression by a factor of a. The factor $1/\sqrt{a}$ is not essential in the theory, but it is very convenient because it preserve L^2 norms, *i.e.*, $\|\psi\|_2 = \left\|\psi_{a,b}\right\|_2$.

The Continuous Wavelet Transform can then be defined as a standar inner product:

Definition 40. *The **Continuous Wavelet Transform (WT)** of a given function $f \in L^2(\mathbb{R})$ with respect to a mother wavelet ψ is given by:*

$$W_\psi[f](a,b) = \langle f, \psi_{a,b}\rangle = \int_{\mathbb{R}} f(t) \cdot \psi_{a,b}^*(t)\, dt \qquad (a \in \mathbb{R}_+^*, b \in \mathbb{R}),$$

which is the standard inner product in a Complex Hilbert Space.

Like in the case of FT, it would be desirable to invert the transform to recover some information about the function in the time domain. In the STFT, the inversion is also possible for certain functions like in the FT, which is intuitive, since the STFT is itself a FT. For the WT, however, the situation is more delicate and the result is given in the next theorem:

Theorem 26. *Let $\psi \in L^2(\mathbb{R}) \cap L^1(\mathbb{R})$ be a real valued function that satisfies the following requirement, know as the admissibility condition:*

$$c_\psi = \int_0^\infty \frac{|\mathcal{F}[\psi](\omega)|^2}{\omega}\, d\omega < \infty.$$

Then, for any $f \in L^2(\mathbb{R})$,

$$\|f\|_2 = \sqrt{c_\psi}\left(\int_{(a,b)\in(0,\infty)\times\mathbb{R}} |W_\psi[f](\omega)|^2 \frac{da}{a^2}db\right)^{1/2}.$$

Moreover, if $f \in L^p(\mathbb{R})$, $1 < p < \infty$, and the integral is interpreted in a distributional sense, then:

$$f(t) = \frac{1}{c_\psi}\left(\int_{(a,b)\in(0,\infty)\times\mathbb{R}} W[f](a,b)\psi_{a,b}(t)\frac{da}{a^2}db\right).$$

Example 30. *Consider the Haar Wavelet:*

$$\psi(t) = \begin{cases} 1, & \textbf{\textit{if}}\ 0 \le t < 1/2; \\ -1, & \textbf{\textit{if}}\ 1/2 \le t < 1; \\ 0, & \textbf{\textit{if}}\ t > 1\ \textbf{\textit{or}}\ t < 0; \end{cases}$$

Then, its Fourier Transform is given by

$$\mathcal{F}[\psi](\omega) = \frac{\left(1 - e^{-\pi i\omega}\right)^2}{2\pi i\omega}.$$

This is in fact admissible.

4.7 Discrete Wavelet Transform And Discretization Of Continuous Wavelet Transform

Just as in the case of Fourier Transforms, one may wish to consider Wavelet Transform of discrete signals. However, in the context of wavelets, the concept of Discrete Wavelet Transforms can mean two different things, both very usual to see in this context. The first one is the **Wavelet Transform Of**

A Discrete Signal and the second is the **Discretization Of The Continuous Wavelet Transform**. Both will be shorten by DWT and will be distinguished by the context and presented next.

Given a signal in the domain of $\mathbb{Z}$, which will be a sequence $f[n]$, its DWT consists of passing this signal through a cascade of filters which, at each level, consists of one high pass filter and one low pass filter. In the first step, the samples pass through a low pass filter with impulse response g, which will result in the following convolution of signals:

$$y[n] = (f * g)[n] = \sum_{k=-\infty}^{\infty} f[k]g[n-k].$$

This same signal is also passed trough a high pass filter h. The outputs of the low pass filters are called the **approximation coefficients**, whilst the coefficients of the high pass filters are called the **detail coefficients**. These two filters must constitute a **quadrature mirror filter** of one another, *i.e.*, the magnitude response is the mirror image around $\pi/2$ of the other filter. This process continues with the outputs of the low pass filters (the approximation coefficients). This tree of filters is know as a **filter bank**.

At each step, which continues only with the approximation coefficients, half of the samples are discarded, which can be done by the Shannon-Nyquist Sampling Theorem. The approximation coefficients are subsampled by 2 and are further passed again by a low pass filter g and a high pass filter h with half of the cut-off frequency of the previous ones, *i.e.*:

$$\begin{cases} y_{\text{low}} = \displaystyle\sum_{k=-\infty}^{\infty} f[k]g[2n-k]; \\ y_{\text{high}} = \displaystyle\sum_{k=-\infty}^{\infty} f[k]h[2n-k] \end{cases}$$

At each step, this decomposition halves the time resolution because only half of each filter output characterises the signal. On the other hand, each output has half the frequency band of the previous and thus the frequency resolution is doubled.

All this process of decomposition is illustrated in Figure 4.10 bellow.

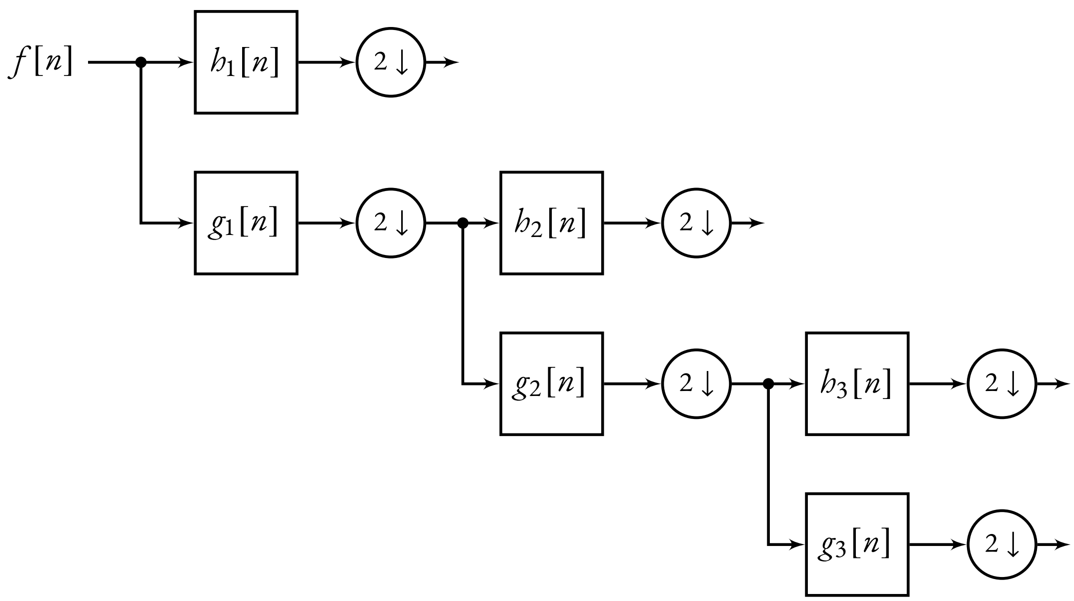

Figure 4.10: Filter bank for Discrete Wavelet Transform.

Denoting the **Subsampling Operator** by $\downarrow$:

$$(y \downarrow k)[n] = y[kn],$$

the sums for y_{low} and y_{high} expressed before can be written in a more succinct form:

$$\begin{cases} y_{\text{low}} = (f * g) \downarrow 2; \\ y_{\text{high}} = (f * h) \downarrow 2 \end{cases}$$

The other usual way to see the discrete wavelet transforms is the discretization of the Continuous Wavelet Transforms. This is usually done to avoid redundancy of information and optimize calculations for computational performance. In the case of Haar wavelets for example, there is much redundancy information in $W[f](a,b)$. The wavelets in the family with $a = 1$, for example, constitute a orthonormal set in $L^2(\mathbb{R})$:

$$\{\psi_{1,m} : m \in \mathbb{Z}\} = \{\mathbb{1}_{[m,m+1/2)} - \mathbb{1}_{[m+1/2,m+1)} : m \in \mathbb{Z}\}.$$

This is not a basis for the square-integrable function. It is in fact a basis for the piece-wise constant functions on a interval such that the constant parts are the intervals defined by these functions (endpoints at adjacent half integers).

Haar, however, in his original work (see [14]), showed that the following set is indeed a orthonormal (Hilbert) basis for $L^2(\mathbb{R})$:

$$\{\psi_{2^n,2^n m} : n, m \in \mathbb{Z}\}.$$

Hence, every f in $L^2(\mathbb{R})$ can be expressed as a infinite linear combination of the projections against these wavelets in the family:

$$f = \sum_{n,m\in\mathbb{Z}} \langle f, \psi_{2^n,2^n m}\rangle \psi_{2^n,2^n m} = \sum_{n,m\in\mathbb{Z}} W[f](2^n, 2^n m)\psi_{2^n,2^n m}.$$

However, other works showed later that other admissible wavelets ψ can be used in place of the Haar wavelet and still form a Hilbert basis for $L^2(\mathbb{R})$. This is convenient for example when one wants smoothness.

A

Riesz-Thorin's Interpolation Theorem

A.1 Riesz-Thorin's Interpolation Theorem

During this section, the central aim is to explore the Riesz-Thorin's Interpolation Theorem and show some applications, particularly, Young's Inequality previously mentioned in the last section. The proof for this theorem uses a complex analysis theorem called the *Hadamard Three-Lines Theorem*, which is the starting point to this section.

In the proof of this Hadamard's Theorem, the *Maximum Principle* will be used, that basically says that if $f(z)$ is a function of complex variable continuous in a domain[†] S and holomorphic (*i.e.*, analytic) in the interior of this domain, then its maximum will be attained in the boundary of this domain.

Theorem 27 (Hadamard Three-Lines)**.** *Let $F(z)$ be a continuous function and bounded in the domain*

$$S = \{z = x + iy : 0 \leq x \leq 1\},$$

such that this function is also holomorphic in the interior of S. If for all $y \in \mathbb{R}$ it holds that $|F(iy)| \leq M_0$ e $|F(1 + iy)| \leq M_1$, then, for all $z = x + iy \in S$, one has that $|F(x + iy)| \leq M_0^{1-x} M_1^x$.

Proof. Suppose that $M_0, M_1 > 0$ (which is the case of interest). Indeed, it suffices to prove for $M_0 = M_1 = 1$ taking the auxiliary function $G(z) = \frac{F(z)}{M_0^{1-z} M_1^z}$. Hence, $|G(iy)| \leq 1$ and $|G(1+iy)| \leq 1$. Then, the goal is to prove that $|G(z)| \leq 1$ for all $z \in S$.

Start defining $G_n(z) = G(z) e^{\frac{(z^2-1)}{n}}$ for $n \in \mathbb{Z}_+$. Now, notice that

$$|G_n(z)| = |G(x + iy)| e^{-\frac{y^2}{n}} \cdot e^{\frac{x^2-1}{n}} \leq |G(x + iy)| e^{-\frac{y^2}{n}},$$

since $0 \leq x \leq 1$. Therefore, for any $n \in \mathbb{Z}_+$, the function G_n converges to zero 0 when $|y| \longrightarrow \infty$ uniformly in $0 \leq x \leq 1$.

Also, observe that $|G_n(iy)| = |G(iy)| e^{-\frac{y^2+1}{n}} \leq 1$ and, therefore, it can be concluded in an analogous way that $|G_n(1 + iy)| = |G(1 + iy)| e^{-\frac{y^2}{n}} \leq 1$. Since the function G_n converges to 0, there must exist a $|y_0|$ such that $|y| \geq |y_0| \rightarrow |G_n(x + iy_0)| \leq 1$. Hence, it is possible to see that in the boundary of the rectangle with vertices $(0, iy_0)$; $(1, iy_0)$; $(1, -iy_0)$; $(0, -iy_0)$, the function is bounded by 1. Therefore,

[†] Here domain means only the domain of the function, *i.e.*, the set in which the function is defined. In Hadamard Three-Lines Theorem indeed, this domain will be a strip as will be clear when it is stated.

by the maximum principle, this same limit applies in the interior of the rectangle. Hence, for every $n \in \mathbb{Z}_+$, the result holds. As taking $n \to \infty$, it converges to $G(z)$, the desired result is valid for $G(z)$, concluding this proof. □

Definition 41. *Let $T : L^p \to L^q$ be a bounded linear operator. the norm of this operator is defined as*

$$||T||p, q = \sup \frac{||Tf||_q}{||f||_p} = \sup_{||f||_p=1} ||Tf||_q.$$

Observation 8. *Let $h \in L^q$, and q' the conjugate of q. Then, it is possible to write $||h||q = \sup ||g||_{q'} = 1|\langle h,g\rangle|$ where $\langle h,g\rangle = \int h(y)g(y)dy$.*

The essence of the Riesz-Thorin's Interpolation Theorem it to make possible to define a convergence region and a bound for a operator, verifying only for certain pairs (p, q) where the operator $T : L^p \to L^q$ is defined and bounded. That being said, now the main theorem of this section can be stated and proved.

Theorem 28 (Riesz-Thorin's Interpolation)**.** *Let T be a linear operator bounded in the spaces:*

$$T : L^{p_0} \to L^{q_0}$$

and

$$T : L^{p_1} \to L^{q_1},$$

satisfying $||T||p_0, q_0 \leq M_0$ and $||T||p_1, q_1 \leq M_1$. Define

$$\frac{1}{p_\theta} = \frac{1-\theta}{p_0} + \frac{\theta}{p_1}$$

e

$$\frac{1}{q_\theta} = \frac{1-\theta}{q_0} + \frac{\theta}{q_1}.$$

Then, $||T||p\theta, q_\theta \leq M_0^{1-\theta} \cdot M_1^\theta$ *for* $0 \leq \theta \leq 1$.

Proof. Start noticing that $p_\theta, q'\theta < \infty$. Therefore, it is known that the class of simple functions with compact support is dense in $L^{p\theta}$ and $L^{q'_\theta}$, allowing the proof to be made only for these class of functions and the general case follows by density.

Consider then $f \in L^{p_\theta}$ and $g \in L^{q'\theta}$, both with unitary norm and being of the form $f = \sum_{j=1}^n a_j \mathbb{1}_{E_j}$ and $g = \sum_{k=1}^m b_k \mathbb{1}_{A_k}$, respectively. Let also $p(z)$ and $q'(z)$ be defined in the following way:

$$\begin{cases} \dfrac{1}{p(z)} = \dfrac{1-z}{p_0} + \dfrac{z}{p_1} \\ \dfrac{1}{q'(z)} = \dfrac{1-z}{q'_0} + \dfrac{z}{q'_1}. \end{cases}$$

Define $\varphi(z) = \sum_{j=1}^n |a_j|^{\frac{p_\theta}{p(z)}} \cdot e^{i \cdot arg(a_j)} \cdot \mathbb{1}_{E_j}$ and $\phi(z) = \sum_{k=1}^m |b_k|^{\frac{q'\theta}{q'(z)}} \cdot e^{i \cdot arg(b_k)} \cdot \mathbb{1}_{A_k}$. Notice that both of these functions are continuous in the domain S defined in the Hadamard Three-Lines Theorem. Moreover, being exponentials, they are also analytic in the interior of this domain and, therefore, the function $F(z) = \langle T\varphi, \phi \rangle$ also has these properties. Now the goal is to verify that $\varphi(iy) \in L^{p_0}$ and $\varphi(1+iy) \in L^{p_1}$, and, in an analogous way, that $\phi(iy) \in L^{q'0}$ and $\phi(1+iy) \in L^{q'_1}$. To show that, notice that $|\varphi(iy)| = \sum j = 1^n |a_j|^{\frac{p_\theta}{p_0}} \cdot \mathbb{1}_{E_j} = |f|^{\frac{p_\theta}{p_0}}$. Hence, $||\varphi(iy)||p_0 = ||f||p_\theta^{\frac{p_\theta}{p_0}} = 1$ because the norm of f is unitary in L^{p_θ}. The other claims can be verified in an entirely analogous way.

Due to the characterization of the norm given in the Observation 8, notice that $|F(iy)| = |\langle T\varphi(iy), \phi(iy) \rangle| \leq ||T||p_0, q_0 \leq M_0$ and also $|F(1+iy)| = |\langle T\varphi(1+iy), \phi(1+iy) \rangle| \leq ||T||p_1, q_1 \leq M_1$, and, therefore, Theorem 27 can be applied. Since for $z = \theta$ one has that $\varphi(\theta) = f$ and $\phi(\theta) = g$, this leads to $|F(\theta)| = |\langle Tf, g \rangle| \leq M_0^{1-\theta} M_1^\theta$. Since this is valid for any $f \in L^{p_\theta}$ and $g \in L^{q'\theta}$ with unitary norm, this means that $||T||p_\theta, q'_\theta \leq M_0^{1-\theta} M_1^\theta$, concluding the proof. □

A.2 Applications

Now two very well known results that can be proved using the Riesz-Thorin's Interpolation Theorem will be presented. The first one is known as the Hausdorff-Young Inequality and it is very important in the development of the theory of Fourier Transforms.

Lemma 9 (Hausdorff-Young Inequality). *Let $f \in L^p(\mathbb{R}^n)$, $1 \leq p \leq 2$. Then, $\hat{f} \in L^{p'}(\mathbb{R}^n)$ (where p' is conjugate to p, i.e., $\frac{1}{p} + \frac{1}{p'} = 1$) and $||\hat{f}||_{p'} \leq ||f||_p$, where $\hat{f}$ represents the Fourier Transform of f.*

Proof. First, recalling the theory of Fourier Transforms, it is already known that indeed it concerns a linear operator $T : L^1 \to L^\infty$, such that $T(f) = \hat{f}$ and also $T : L^2 \to L^2$. Moreover, it is also known that $||\hat{f}||\infty \leq ||f||_1$, hence $||T||1, \infty \leq 1$, and, in L^2, one has a unitary map and, therefore, one also has $||T||_{2,2} \leq 1$.

Thus, due to Riesz-Thorin's Interpolation Theorem there must exist p, q with

$$\frac{1}{p} = 1 - \theta + \frac{\theta}{2}$$

and

$$\frac{1}{q} = 0 + \frac{\theta}{2},$$

such that $||T||p, q \leq 1$. It is easy to see from the definition of p, q that $1 \leq p \leq 2$ and $\frac{1}{p} + \frac{1}{q} = 1$, hence $q = p'$. Since $\frac{||\hat{f}||p'}{||f||p} \leq ||T||p, p'$, the fact that $T_{p,p'}$ is bounded implies the desired inequality. □

Finally, next the Young's Inequality for convolutions will be presented. Its proof concerns an ingenious use of the Riesz-Thorin's Interpolation Theorem.

Lemma 10 (Young's Inequality). *Let $f \in L^p(\mathbb{R}^n)$ and $g \in L^q(\mathbb{R}^n)$, with $1 \leq p, q \leq \infty$, and $\frac{1}{p} + \frac{1}{q} \geq 1$.*

Then, $f * g \in L^r(\mathbb{R}^n)$*, where* $\frac{1}{r} + 1 = \frac{1}{p} + \frac{1}{q}$ *with*

$$||f * g||_r \leq ||f||_p \cdot ||g||_q.$$

Proof. Start fixing a q and a $g \in L^q$, and defining a operator T that makes the convolution of a certain function with g. Now the idea is to verify that this operator is of the form $(1, q)$ and (q', ∞).

To do that, let $f \in L^1$, and consider the following calculation:

$$\begin{aligned} ||f * g||_q &= \left(\int \left|\int g(x-y)f(y)dy\right|^q dx\right)^{\frac{1}{q}} \\ &\leq \int f(y) \left(\int |g(x-y)|^q dx\right)^{\frac{1}{q}} dy \\ &\leq ||f||_1 \cdot ||g||_q, \end{aligned}$$

where the Minkowski inequality for integrals was used. Hence, it is possible to see that $||T||1, q \leq ||g||_q$. On the other hand, if $f \in L^{q'}$, the Holder Inequality can be applied to conclude that $||f*g||\infty \leq ||f||q'||g||_q$, which implies that $||T||q', \infty \leq ||g||_q$.

That being said, due to Riesz-Thorin's Interpolation Theorem, T must be of the type (p, r), satisfying the following conditions:

$$\begin{aligned} \frac{1}{p} &= 1 - \theta + \frac{\theta}{q'} \\ \frac{1}{r} &= \frac{1-\theta}{q}. \end{aligned}$$

Subtracting the equations above the desired relation follows:

$$\frac{1}{r} + 1 = \frac{1}{p} + \frac{1}{q}.$$

□

Recall that the Fourier Transform was initially defined for functions $f \in L^1$ as, the Fourier Transform $\hat{f} = \mathcal{F}[f]$ as

$$\mathcal{F}[f](\xi) = \int_{\mathbb{R}^n} f(x) e^{-i\xi \cdot x}\, dx.$$

Then, in order to define the transform in L^2, the central idea was to use the density of the Schwartz' Class $\mathcal{S}$, as $\mathcal{S}(\mathbb{R}^n)$ is dense in $L^2(\mathbb{R}^n)$. Hence, given $f \in L^2(\mathbb{R}^n)$ the procedure was to take a sequence of functions in $\mathcal{S}(\mathbb{R}^n)$ (since $\mathcal{S}(\mathbb{R}^n)$ is dense in $L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$), converging to f in the L^2 norm, *i.e.*, $\|f - f_n\|_2 \xrightarrow{n \longrightarrow \infty} 0$. Now, since the Fourier Transform is an isometry in L^2, then $\|g\|_2 = \|\hat{g}\|_2$ for $g \in \mathcal{S}(\mathbb{R}^n)$. Therefore, the sequence defined by $\hat{f}_n$ is a Cauchy sequence in $L^2(\mathbb{R}^n)$, and, hence, there must exist a unique $\hat{f}$ such that $\hat{f}_n \xrightarrow{n \longrightarrow \infty} \hat{f}$. The Fourier Transform of f was defined as this limit $\hat{f}$.

In the case where $1 \leq p \leq 2$, the same idea can be applied because the Schwartz Class is actually dense in $L^p(\mathbb{R}^n)$ for all $1 \leq p < \infty$ †. Therefore, given $1 < p < 2$, it is possible to define the Fourier Transform for functions in $\mathcal{S}(\mathbb{R}^n) \cap L^p(\mathbb{R}^n)$. Therefore one can take in the same way a sequence of functions in $\mathcal{S}(\mathbb{R}^n)$ (since $\mathcal{S}(\mathbb{R}^n)$ is dense in $L^p(\mathbb{R}^n) \cap L^{p'}(\mathbb{R}^n)$), converging to f in the L^p norm, *i.e.*, $\|f - f_n\|_p \xrightarrow{n \longrightarrow \infty} 0$ and the p' is the conjugate of p (Hausdorff-Young Inequality). Now, it is known that $\|f - f_n\|_p \xrightarrow{n \longrightarrow \infty} 0$ (in particular it is a Cauchy sequence), due to Riesz-Thorin Theorems, there will be induced a sequence $\hat{f}_n$ in $L^{p'}$, and since $\mathcal{F}$ is a bounded linear operator, this theorem implies that the induced sequence will also be a Cauchy sequence in $L^{p'}$. Finally, since $L^{p'}$ is complete (indeed a Banach space), this sequence will converge to a unique limit $\hat{f}$, and the Fourier Transform of $f \in L^p$ will be defined as this $\hat{f} \in L^{p'}$.

† For L^∞ this is not true as the constant function $F \equiv 1$ is in L^∞ and can not be approximated by functions in $\mathcal{S}$, since they are of slow growth and therefore the L^∞ norm will not go to zero outside any interval that is not the entire space.

B

Codes To Generate The Figures Presented In The Signal Processing Subsection

In this appendix, the codes to generate the figures of the signal, the STFT's and the Wavelet Transform graphics will be presented.

B.1 Code For The Graphic of The Chirp Signal With High Frequency Interjections

The code presented bellow plots the chirp signal with high frequency interjections using Python language.

```
from pywt import cwt, frequency2scale, ContinuousWavelet
import numpy as np
from scipy.signal import stft
import matplotlib.pyplot as plt

plt.style.use("ggplot")

dt = 0.01
time = np.arange(0, 200, dt)

signal = np.sin(time * (100*(time//25*dt)) + 50)
plt.plot(signal)

f, t, short_time = stft(signal, fs=1/dt)
plt.imshow(np.abs(short_time))
plt.show()

frequencies = np.linspace(100, 25, 2000) * dt
scale = frequency2scale('cmor1.5-1.0', frequencies)
wav = ContinuousWavelet('cmor1.5-1.0')

coefs, f = cwt(signal, scale, wav)
```

```
plt.imshow(np.abs(coefs))
plt.show()
```

B.2 Code For The Graphics of two STFT's and a WT of The Chirp Signal With High Frequency Interjections

This other code presented next provides and plots two STFT's and a WT for the chirp signal with high frequency interjections also using Python language.

```
import pywt
import numpy as np
from scipy.signal import stft
import matplotlib.pyplot as plt

cmap="magma"

fs = 44100
length = 10e-3
n_samples = np.int(length*fs)

time = np.linspace(0, length, n_samples)
signal = 2*np.cos(2*np.pi*500*time * (time // (50/fs)))
signal[85:88] = 0
signal[300:304] = 0.5

signal = signal - signal.mean()

plt.style.use("ggplot")
```

```python
plt.plot(time, signal)
plt.xlabel("Time [s]")
plt.ylabel("Amplitude")
plt.title("Time Domain Signal")
plt.show()

plt.style.use("default")

f, t, short_time = stft(signal, fs=fs, nperseg=8)
plt.imshow(abs(short_time), cmap=cmap, extent=[t[0], t[-1], f[-1], f[0]],
           aspect="auto", interpolation="bilinear")
plt.xlabel("Time [s]")
plt.ylabel("Frequency [Hz]")
plt.title("Short Time Fourier Transform with Window Size 8")
plt.colorbar()
plt.show()

f, t, short_time = stft(signal, fs=fs, nperseg=256)
plt.imshow(abs(short_time), cmap=cmap, extent=[t[0], t[-1], f[-1], f[0]],
           aspect="auto", interpolation="bilinear")
plt.xlabel("Time [s]")
plt.ylabel("Frequency [Hz]")
plt.title("Short Time Fourier Transform with Window Size 256")
plt.colorbar()
plt.show()

scales = np.arange(1, 21, 1)

```

```
coef, freqs = pywt.cwt(signal, scales, "mexh")
plt.figure(figsize=(15, 10))
plt.imshow(abs(coef), cmap=cmap, extent=[0, 10e-3, 20, 1],
           interpolation="bilinear", aspect="auto")
plt.gca().invert_yaxis()
plt.xticks(np.arange(0, n_samples/fs, 2*n_samples/(20*fs)))
plt.yticks(scales)
plt.colorbar()
plt.xlabel("Time [s]")
plt.ylabel("Scale [s]")
plt.title("Wavelet Transform")
plt.show()
```

References

[1] B. B. H. (Author). *The World According to Wavelets: The Story of a Mathematical Technique in the Making, Second Edition*. A K Peters/CRC Press, 2 edition, 1998.

[2] R. G. Bartle. *The Elements of Integration and Lebesgue Measure*. Wiley Classics Library, 1995.

[3] H. Brezis. *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Universitext. Springer-Verlag New York, 1 edition, 2010.

[4] Brilliant.org. Distributions. `https://brilliant.org/wiki/distributions/`.

[5] H. Carlson. *Lecture Notes on Distributions*. 2011.

[6] E. Carneiro. *Aulas do Curso de Equações Diferenciais Parciais*. Impa.

[7] R. F. Evans, Lawrence Craig; Gariepy. *Measure Theory and Fine Properties of Functions, Revised Edition*. Textbooks in Mathematics. CRC Press, 2015.

[8] G. B. Folland. *Real analysis: modern techniques and their applications*. PAM. Wiley, 2 edition, 1999.

[9] G. B. Folland. *A Course in Abstract Harmonic Analysis*. Textbooks in Mathematics. Chapman and Hall/CRC, 2 edition, 2015.

[10] J. P. Groβmann. Bright side of mathematics: Distribution theory. `https://thebrightsideofmathematics.com/courses/distributions/overview/`.

[11] L. Hörmander. *The analysis of linear partial differential operators*, volume I. Springer-Verlag, 2nd ed edition, 1990.

[12] J. Jost. *Postmodern Analysis*. Springer, 1998.

[13] R. P. Kanwal. *Generalized Functions: Theory and Applications*. Birkhäuser Basel, 3 edition, 2004.

[14] D. E. Knuth. Zur theorie der orthogonalen funktionensysteme. *Mathematische Annalen*, 1910.

[15] B. S. Michael Reed. *Methods of modern mathematical physics 1: Functional Analysis*, volume 1. Academic Press, rev. and enl. ed edition, 1981.

[16] J. Munkres. *Topology*. Prentice Hall, Inc, 2nd ed edition, 2000.

[17] A. L. Neto. *Funções de uma variável complexa*. Projeto Euclides. Impa, 2016.

[18] A. Sasane. *Friendly Approach To Functional Analysis, A*. Essential textbooks in mathematics. World Scientific Publishing Company, 2017.