# WEEK 5: REVIEW OF ANALYSIS (CONTINUED)

## 1. The $p$ Norms on Function Spaces

In this section, we look at a class of norms on functions called the $p$-norms. Although these can be defined for functions on any measure space, an idea that would be introduced in a measure theory course like MAT 473 or MAT 570, we will confine our attention to the three settings most often encountered: ordered $n$-tuples, sequences, and functions on $\mathbb{R}^n$.

The part on the the Schwarz class is adapted from [2].

### 1.1. Equivalent norms.

Before we begin defining these norms, it is worth presenting the concept of equivalent norms, since it is worthwhile to know how these norms might relate to one another:

**Definition 1.1.** Two norms $\|\cdot\|_a$ and $\|\cdot\|_b$ on a vector space $A$ are said to be *equivalent* if there $N, M > 0$ such that for all $x \in A$, we have

$$N \|x\|_a \leq \|x\|_b \leq M \|x\|_a .$$

Equivalent norms define the same topology on the space $A$, i.e., the same sequences converge, the same sets are open, and the same functions are continuous. We leave the proofs as exercises. Since the ideas of convergence, continuity, and openness are the fundamental tools of analysis, the differences between equivalent norms can usually be ignored.

We will assume without proof that all the spaces we present in this section are complete.

### 1.2. Ordered $n$-tuples.

Many of these norms already appeared in the examples, but we define them again in order to give a relatively complete discussion, and see the parallels between the simple case of ordered $n$-tuples and norms on infinite-dimensional spaces.

1.2.1. *The Uniform Norm.* The uniform or max norm on $\mathbb{R}^n$ and $\mathbb{C}^n$ is given by

$$\|x\|_u = \max_{1\leq j\leq n} |x_j| ,$$

where the $x_j$'s are the coordinates of $x$. This norm is also sometimes called the 'infinity norm' and denoted $\|cdot\|_\infty$. The max norm gives nice coordinate-wise estimates on a vector. It is used on $\mathbb{R}^n$ and $\mathbb{C}^n$ primarily because it is related to the standard Euclidean norm, which allows us to show limits, continuity, and convergence coordinate-wise. For example, the easies way to prove that $\mathbb{R}^n$ and $\mathbb{C}^n$ are complete is by using the max norm combined with the fact that $\mathbb{R}$ is.

1.2.2. *The p-norms.* For $1 \leq p < \infty$, the $p$-norms on $\mathbb{R}^n$ and on $\mathbb{C}^n$ are defined by

$$\|x\|_p = \left( \sum_{j=1}^{n} |x_j|^p \right)^{1/p} .$$

The 1-norm, which also appeared as an example earlier, is sometimes called 'taxicab' norm, since it gives the distance one would have to travel along a grid to get from the origin to the point in $n$-space. It is worth pointing out that for even $p$, we can leave the absolute values out in the definition of the $p$-norm on $\mathbb{R}^n$, but not in $\mathbb{C}^n$. Although it is clear from linearity of the sum that the taxicab norm is a norm and we know from Cauchy-Schwarz that the 2-norm is a norm, the triangle inequality is not quite trivial to check for the other $p$ norms; we will prove it after we have defined the analogous norms on sequences and on continuous function spaces.

1.2.3. *The 2-norm and the Inner Product.* The 2-norm, often called the standard Euclidean norm, is used almost exclusively on $n$-space, mainly because it makes the best geometric sense due to the fact that it is associated with the inner product on $n$-space, the 'dot' product

$$x \cdot y = \sum_{j=1}^{n} x_j \bar{y}_j .$$

We saw this inner product earlier in the examples. The 2-norm is the only $p$-norm associated with an inner product.

1.2.4. *Relationships between these norms.* A useful result from finite-dimensional normed linear algebra, which we will not prove here, is that *any* two norms on a finite-dimensional space are equivalent. Thus, to some extent, the section below regarding $p$-norms on $\mathbb{C}^n$ is unnecessary, other than the fact that in some contexts one norm will be more convenient than another, and it is nice to know that we can interchange them. We use one manifestation of this result, for example, when we argue that a function from $\mathbb{R}^n$ to $\mathbb{R}^m$ is continuous if all of the coordinate maps are. In this case, we are using the fact that the Euclidean 2-norm is equivalent to the max norm.

No such result holds in infinite-dimensional vector spaces, so in the sequence spaces and function spaces that we describe below, the norms are all distinct, and defined over distinct vector spaces.

1.3. Sequences of Complex Numbers.

Before defining these spaces and norms, we remark that although we give the definitions for bi-infinite sequences, since these are the sequences which have the deepest connections with Fourier analysis, the definitions make perfectly good sense for ordinary sequences on the natural numbers.

In the context of ordered $n$-tuples, we were able to define the norms without any caveats about the spaces on which they were defined, because the formulas were clearly well-defined, using only finite sums. Because sequences are infinite-dimensional, there is usually an issue of convergence of the norm, and we have to restrict our attention only to certain sequences in order to have normed vector spaces.

1.3.1. *The Sup Norm and $\ell^\infty(\mathbb{Z})$.* We define the space $\ell^\infty(\mathbb{Z})$ to be the space of bounded sequences of complex numbers, that is, those whose images are bounded subsets of the complex plane, and we define the sup or uniform norm, also called the $\infty$ norm, on $\ell^\infty(\mathbb{Z})$ by [1]

$$\|\{a_n\}\|_\infty = \sup_{j\in\mathbb{N}} |a_j| .$$

The '$\infty$' in the subscript is often replaced by '$u$' or by 'sup.' Note that, in order for $\|\cdot\|_\infty$ to be a norm, we had to restrict our attention to bounded sequences. Otherwise, the norm could take the value of $\infty$, which is not allowed for a norm. In fact, $\ell^\infty(\mathbb{Z})$ is the very largest subspace of the space of all sequences on which $\|\cdot\|_\infty$ is a norm, since the condition of boundedness is both necessary and sufficient for $\|\cdot\|_\infty$ to be finite on a particular sequence. Technically, we need to check that $\ell^\infty(\mathbb{Z})$ is closed under scalar multiplication and addition, but this follows easily by considering the sequences pointwise.

1.3.2. *The $p$-norms.* The $p$-norms, $1 \le p < \infty$, are defined on sequences of complex numbers by

$$\|\{a_n\}\|_p = \left(\sum_{j\in\mathbb{Z}} |a_j|^p\right)^{1/p} .$$

For each $p$, the space of complex-valued sequences for which the above sum converges is denoted $\ell^p(\mathbb{Z})$. Once again, these are the largest subspaces of the space of all complex sequences on which the corresponding norms take only finite values. Note that any sequence in $\ell^1$ defines an absolutely convergent series.

Again, we should technically check that $\ell^p(\mathbb{Z})$ is closed under addition and scalar multiplication, but we will leave this proof for after we define the norms for functions on continuous domains.

1.3.3. *The 2-norm and the Inner Product.* As with $n$-tuples, the 2-norm is associated with an inner product,

$$\langle \{a_n\}, \{b_n\} \rangle = \sum_{j=1}^{\infty} a_j \bar{b}_j .$$

We should check that this series converges whenever $\{a_n\}, \{b_n\} \in \ell^2(\mathbb{Z})$, but once again we will put off the proof until after we cover the analogous norm and inner product in continuous spaces.

1.3.4. *Relationships between these norms.* Although none of the norms above are equivalent, they do have a convenient ordering, which can be summarized by two facts:

**Proposition 1.2.** *If $1 \le p \le q \le \infty$, then $\ell^p(\mathbb{Z}) \subset \ell^q(\mathbb{Z})$.*

---

[1] Some authors use $\|\cdot\|_{\ell^\infty}$ and $\|\cdot\|_{\ell^p}$ in place of $\|\cdot\|_\infty$ and $\|\cdot\|_p$ when dealing with sequences, in order to emphasize that the norm is defined by a sum, rather than an integral like that defining the $L^p$ norms for functions on $\mathbb{R}^n$. However, from a measure-theoretic point of view, a sum is simply an integral with respect to counting measure, so the distinction is rather minor. We will explicitly notate the space on which we take the norm when important.

*Proof.* If $q = \infty$, this is obvious, because an unbounded sequence cannot possibly have a finite sum. Hence, let $a_n \in \ell^q$. There is $N \in \mathbb{N}$ such that for $|n| \geq N$, $|a_n| < 1$. The finitely many terms $a_1, ..., a_N$ have no bearing on convergence of the norm, and since $p \geq q$, it follows that $|a_n|^p \leq |a_n|^q$ for $|n| \geq N$. Hence,

$$\sum_{|n| \geq N} |a_n|^p \leq \sum_{|n| \geq N} |a_n|^q .$$

□

It is important to realize that the $n$ in the proposition below refers to the indexing of the sequences $s_n$, and NOT to the indexing of the coordinates within those sequences.

**Proposition 1.3.** *If $1 \leq p \leq q \leq \infty$, and a sequence of sequences $\{x_j^n\} \in \ell^p(\mathbb{Z})$ converges in $\ell^p$ to $\{x_j\} \in \ell^p(\mathbb{Z})$, then $x_j^n$ converges to $x_j$ in $\ell^q$.*

*Proof.* It follows from the proposition above that each $x_j^n$ and $x_j$ are all $\ell^q$. The mechanics of the proof are similar to those in the previous lemma, except that we look at the differences $\{x_j - x_j^n\}$. We leave the details to the reader. □

This leads naturally to the concept of embedding:

**Definition 1.4.** Let $X$ and $Y$ be metric spaces. We say $X$ is *embedded* in $Y$ if there is an injective map $\phi : X \to Y$ such that if $x_n \to x$ in $X$, then $\phi(x_n) \to \phi(x)$ in $Y$.

The two propositions above show that for $p \leq q$, $\ell^p$ is embedded in $\ell^q$ by the inclusion map $\phi(x) = x$. Not all embeddings have to be inclusive; for example $\mathbb{R}^n$ is embedded in $\mathbb{R}^m$ for any $n \leq m$ (any injective linear map suffices to prove this) even though it is not a subset of $\mathbb{R}^m$.

### 1.4. Functions on $\mathbb{R}^n$.

We finally move to the spaces which will be of most interest in Fourier analysis—spaces of functions on continuous domains.

1.4.1. *The Uniform Norm.* On the set of all bounded functions $f : R^n \to \mathbb{C}$ we define the sup norm or uniform norm, denoted $\|\cdot\|_u$ by

$$\|f\|_u = \sup_{x \in \mathbb{R}^n} |f(x)| .$$

Note that on the vector space of bounded functions, this norm is well-defined.

This norm is generally used on the set of continuous functions, for two reasons. One is that—if we restrict our interest to continuous functions on a compact domain—convergence under this norm, which is equivalent to uniform convergence (usually discussed in an advanced calculus class), implies convergence under any of the $L^p$ norms, which we define below. The other is that the continuous functions are complete under this norm, an important fact whose proof we leave to the exercises.

1.4.2. *The p-Norms.* Although we were able to use sums and series to define norms on $\mathbb{C}^n$ and on complex sequences, a discrete sum cannot adequately describe the behavior of a function on $\mathbb{R}^n$, which is defined on an uncountable number of points. Instead, we use integrals to define the $p$-norms for functions on $\mathbb{R}^n$.

For measurable functions $f : \mathbb{R}^n \to \mathbb{C}$, we define the $L^p$ norm of $f$, denoted $\|f\|_{L^p}$ or $\|f\|_p$, by

$$\|f\|_p = \Big( \int_{\mathbb{R}^n} |f(x)|^p \, dx \Big)^{1/p}.$$

The space of functions for which the above integral is finite is denoted $L^p(\mathbb{R}^n)$.

As before, the scalar homogeneity of these norms should be fairly clear; we will prove the triangle inequality shortly. But the reader should be struck by the fact that, unless we require the functions to be continuous, this norm fails positive definiteness. A function which is nonzero at a single point, or even on a large number of points, so long as they carry no $n$-dimensional mass, will have integral zero even though it is not the zero function. This problem can be overcome by adopting the convention that two functions are to be identified if they are equal except on a set of measure zero, that is, equal *almost everywhere.* To be more precise, we can define an equivalence relation $\sim$ on the set of measurable functions by $f \sim g$ if $f - g = 0$ almost everywhere. We leave the proof that $\sim$ is an equivalence relation to the exercise.

Since $\sim$ is an equivalence relation on the set of $L^p$ functions, its equivalence classes partition that set, so we can associate any particular function with a particular equivalence class. And recall that sets of measure zero have no effect on integration properties. This means all the elements in an equivalence class will have the same norm, behave the same with respect to any integral measure or integral transform, and their difference will have norm zero, which means that they should be considered the same function. Indeed, if we view $L^p(\mathbb{R}^n)$ as technically being the set of all such equivalence classes, then we will consider any such two $f$ and $g$ where $f \sim g$ to be the same function.

Now the question is whether this approach is practical in applications. In some contexts, the answer could be no. If the application we are looking at is very sensitive to (or entirely dependent on) the values a function takes on a finite or countable set, this approach could get us into trouble. For example, if we are sampling an unknown function on a countable set, say the integers, to get a general picture of the function, this approach would be troublesome because the integers have measure zero.

However, in most realistic applications, our sample would not be of the values a function takes at a given point, but rather an average in a small neighborhood of a given point. Many chemistry and physics measurements, for example, would be highly irregular if it were possible to take the measurement instantaneously at a given point, because extremely high-frequency white noise affects many such measurements. But in reality, the measurements take some small window of time, and are taken over some small region in space rather than a particular point. In effect, these measurements are small integral averages, which can be mathematically modeled as convolutions, as we shall see in Chapter 5. In other applications as well, we can usually think of measurements being spread out over some interval of time and space, or whatever dimensions apply. In this case, ignoring the possibility that functions might differ on sets of measure zero makes perfect sense, since such functions are essentially indistinguishable.

Another way to look at the same issue is probabilistically. Suppose, in some application, a measurement really *is* dependent on the values a function takes at particular points. Since measurements are never purely exact, there is likely to be some randomness involved in determining exactly where the measurement is taken. For example, a scientist might want to measure the value of a function at 0.9, but in reality there will be a narrow area around 0.9 in which his measurement might actually fall. If the random variable representing where the measurement is taken has a continuous cumulative distribution function (for example, if it is a normal random variable centered at 0.9, no matter how sharply peaked), then the probability of the measurement being taken at any point in a set of measure zero will be zero. Once again, if our measurements are taken at exact points, but there is any randomness in the measurements, then functions of the same equivalence class are indistinguishable with probability 1.

In applications, we usually ignore the fact that $L^p(\mathbb{R}^n)$ is a set of equivalence classes, and treat its elements as though they were functions. As long as we are only using integral operations on them, this is not a problem. This restriction is less problematic than it may seem, because many tools have been developed to work with $L^p$ spaces. For example, the emphasis on integral operations in $L^p$ spaces is the primary reason for the development of weak derivatives, which are integral-based derivatives which are not sensitive to behavior on sets of measure zero.

In applications, the most common $p$ norms by far are the 1-norm, important largely because many integral operators are only sure to converge for $L^1$ functions, and the 2-norm, because (a) the 2-norm comes from a useful inner product and (b) in many physical applications, the 2-norm—also called the square mean—can be thought of as the 'energy' represented in a given function, so that $L^2$ functions are functions with finite energy, which of course is true of any phenomenon in the real world.

*Remark* 1.5. It is worth noting that the primary reason for the development of the Lebesgue integral is to allow these spaces to be complete; the Riemann integrable and continuous subspaces of the $L^p$ spaces are not complete. In some ways, the Lebesgue integrable functions were developed from the continuous functions for the same reason $\mathbb{R}$ was developed from $\mathbb{Q}$.

1.4.3. *The 2-Norm and the Inner Product.* In the case of functions on $\mathbb{R}^n$, the inner product, like the norm, is defined by an integral. Let $f, g \in L^2(\mathbb{R}^n)$. We define

$$\langle f, g\rangle = \int_{\mathbb{R}^n} f(x)\overline{g(x)}dx.$$

This inner product plays a major role in the theory of the Fourier transform.

1.4.4. *The $\infty$-Norm.* One problem with $L^p$ norms is that, except when we look only at continuous functions, they do not have a nice interaction with the uniform norm. The $p$-norms are not at all sensitive to the behavior of functions on a set of measure zero, whereas the uniform norm is sensitive to behavior at even a single point. As a more technical note, since the $p$-norms are defined on equivalence classes modulo sets of measure zero, the uniform norm is not even defined on the same domain as the $p$-norms.

In order to solve these problems, and have a norm defined on equivalence classes as we did with the the $L^p$ norms, we must modify the sup norm using the idea of the essential supremum. We call the resulting norm the $L^\infty$ norm, to indicate its basis in Lebesgue measure.

**Definition 1.6.** Let $f : \mathbb{R} \to \mathbb{R}^n$. We define the *essential supremum* of $f$ by

$$\text{ess sup}\,|f| = \inf\{M : \{x \in \mathbb{R}^n : |f(x)| > M\} \text{ has measure zero}\}.$$

We call the norm defined by the essential supremum the $L^\infty$ norm and denote it by $\|\cdot\|_{L^\infty}$ or $\|\cdot\|_\infty$ .

With this norm, we can use the same equivalence classes used with the $L^p$ norms, and we now have a uniform norm insensitive to behavior on sets of measure zero.

A note on conventions of vocabulary usage is now in order. When we speak of uniform convergence (or, for that matter pointwise convergence) of functions in $L^p$ spaces, we generally mean convergence *almost everywhere*, that is, everywhere except possibly on a set of measure zero. Thus, uniform convergence in $L^p$ spaces means convergence under the $\infty$ norm.

In the case of continuous functions, in any case the only space on which the uniform norm is commonly used, the two norms $\|\cdot\|_u$ and $\|\cdot\|_\infty$ coincide, because a continuous function cannot be nonzero on a set of measure zero unless that set is empty.

### 1.4.5. The $L^p$ Norms on More General Function Spaces.

In this class, the functions on which we use $L^p$ norms are generally defined on $\mathbb{R}^n$, but this need not be the case. The functions could, for example be defined on $\mathbb{C}^n$, or on the set of integers (in fact, if we use a measure on the integers where each point has measure one, we get the $\ell^p$ norms) or on any other space with a Lebesgue measure. One common example of Lebesgue measure spaces other than $\mathbb{R}^n$ on which functions are often defined are probability spaces. Probability is rigorously defined as a measure on some space of possible outcomes. The properties of general measures and measure spaces is a graduate-level topic, and we will not discuss it here here. However, it is worth keeping in mind that $L^p$ spaces can be defined on domains other than $\mathbb{R}^n$.

1.4.6. *Relationships between these norms.* In general, there is no nice relationship between the $L^p$ norms and spaces. This can be seen by looking at the rational functions. Consider the function

$$\left|\frac{1}{x^r}\right| \chi_A(x),$$

where $\chi_A$ is the characteristic function of a set $A$, 1 on $A$ and zero elsewhere. This function is not in $L^p$ for $p \geq r$, but is in $L^p$ for $p < r$. On the other hand, the function

$$\frac{1}{|1+x|^r}, \mathrm{x} \in \mathbb{R}$$

is in $L^p(\mathbb{R})$ for $p > r$, but not for $p \leq r$. Hence, no $L^p$ space is contained in another and no $p$-norm induces a stronger topology than any other. Heuristically, the problem is that for lower values of $p$, $L^p$ is more sensitive to behavior near infinity, while for larger values of $p$, $L^p$ is more sensitive to local singularities where integrals can 'blow up.'

However, in some cases (as with Fourier series), our interest is not in functions defined on all of $\mathbb{R}^n$, but only on functions defined on some bounded subset of $\mathbb{R}^n$, e.g., a bounded interval $I \subset \mathbb{R}$. In this case, we replace the integral over $\mathbb{R}$ in the definition of the norm with an integral only over the interval that we are interested in. In this case, decay at infinity becomes a nonissue, and only singularities matter. Our heuristic view, then, suggests that $L^p(I)$ should be contained in $L^q(I)$ whenever $p > r$, and the topology on $L^p(I)$ should be stronger—that is, convergence under the $p$ norm should imply convergence under the $q$ norm. Our intuition turns out to be correct, and the result holds, although its proof requires use of measure theory so we will omit it.

## 2. Basics of $L^p$ theory

Here we prove that the $p$-spaces are vector spaces, that the $p$-norms satisfy the triangle inequality, and that the inner product converges for vectors for which the 2-norm converges. We begin by presenting, without proof, an inequality that forms the basis for much of $L^p$ theory:

**Proposition 2.1.** *(*Hölder's inequality*) Let $1 \leq p \leq q \leq \infty$, and $\frac{1}{p} + \frac{1}{q} = 1$. Then for any $f \in L^p$ and $f \in L^q$,*

$$\left|\int f(x)g(x)\,dx\right| \leq \|f\|_p \cdot \|g\|_q.$$

The proof of Hölder's inequality is accessible and requires no measure theory, but we omit it because it requires some technical manipulations which are somewhat tedious and not especially instructive. The inequality also applies to sequences and to ordered $n$-tuples.

First, we use Hölder's inequality to show that $L^p$ closed under linear combinations and that the $p$ norm satisfies the triangle inequality, that is, that $L^p$ is a normed vector space:

**Proposition 2.2.** *For $1 \leq p \leq \infty$, $L^p(\mathbb{R}^n)$ is closed under linear combinations, and the $p$ norm satisfies the triangle inequality.*

*Proof.* It is easy to see from linearity of the integral and the form of the norm that $L^p$ is closed under scalar multiplication, so we need only prove the triangle inequality. Let $f, g$ be $L^p$ functions. If $p$ is infinity, the triangle inequality follows easily from the fact that the union of sets of measure zero has measure zero (see exercises), so we may assume $p < \infty$. If $p$ is one, then the result follows from applying the triangle inequality inside the integral and using the fact that for integrable functions $h_1$ and $h_2$, $\int h_1 \leq \int h_2$ if $h_1(x) \leq h_2(x)$ for all $x$. It is also trivial if either $\|f\|_p$ or $\|g\|_p$ is zero. Otherwise, note that by elementary algebra,

$$|f+g|^p \leq (|f| + |g|)\, |f+g|^{p-1}.$$

Next, note that if $q = \frac{1}{1-1/p}$, then $q(p-1) = p$, which implies that $|f+g|^{p-1}$ is in $L^q$. Now, apply Hölder's inequality to get

$$\begin{aligned} \int_{\mathbb{R}^n} |f(x)+g(x)|^p \, dx &\leq \int_{\mathbb{R}^n} (|f(x)| + |g(x)|)\, |f(x)+g(x)|^{p-1} \, dx \\ &\leq \|f\|_p \left\| |f+g|^{p-1} \right\|_q + \|g\|_p \left\| |f+g|^{p-1} \right\|_q \\ &= (\|f\|_p + \|g\|_p) \left( \int_{\mathbb{R}^n} |f(x)+g(x)| \, dx \right)^{1/q}, \end{aligned}$$

which we can rearrange to show that

$$\|f+g\|_p = \left( \int_{\mathbb{R}^n} |f(x)+g(x)| \, dx \right)^{1-1/q} \leq \|f\|_p + \|g\|_p.$$

□

We conclude by presenting, mostly without proof, some facts regarding $L^p$ spaces. A few will be proved in the exercises. Recall the definition of a dense subset of a metric space:

**Definition 2.3.** Let $(A, d)$ be a metric space, and $B \subset A$. We say $B$ is a *dense* subset of $A$ if for each $a \in A$ and each $\epsilon > 0$, there is $b \in B$ such that $d(a, b) < \epsilon$.

**Example 2.4.** A metric space is called *separable* if it has a countable dense subset. So, for example, $\mathbb{R}^n$ is separable because the countable set $\mathbb{Q}^n$ is dense in $\mathbb{R}^n$.

Also recall the definition of the support of a function:

**Definition 2.5.** Let $A$ be a metric space, and let $f : A \to V$ for a vector space $V$ (typically $\mathbb{C}$ or $\mathbb{R}$). The *support* of $f$ is the closure of the set $\{x \in A : f(x) \neq 0\}$.

Before listing some of the spaces that are dense in $L^p$, we should introduce notation for some of the most common classes of functions (all of them are vector spaces, although not all have norms):

- The set of continuous functions on a domain $D$ is denoted $C(D)$.
- The set of functions with a continuous derivative on a domain $D$ is denoted $C^1(D)$, and similarly the set of functions with $k$ continuous derivatives on $D$ is denoted $C^k(D)$, and the set of functions with infinitely many derivatives on $D$ is denoted $C^\infty(D)$.
- The set of continuous functions with bounded support on a domain $D$ is often denoted $C_0(D)$ or $C_c(D)$, and similarly functions of bounded support with derivatives on $D$ are said to be in $C_c^k(D)$. If the functions simply approach zero at infinity, the $c$ is replaced by a 0, e.g. the set of continuous functions which approach zero at $\pm\infty$ is denoted $C_0(\mathbb{R})$.
- The Schwartz space on the real line, denoted $S(\mathbb{R})$, is the set of $C^\infty$ functions on $\mathbb{R}$ for which every derivative also rapidly decreasing (here rapidly decreasing means it approaches zero faster than any rational function). We will look at this space in more detail later.
- The set of infinitely differentiable functions with bounded support is denoted $C_c^\infty$. This set is used to define generalized functions.

Depending on the context, all of the above function spaces can be very convenient to work with (particularly the Schwartz functions). Particularly useful is the fact that functions with each of the above properties are dense in $L^p$ spaces (although, importantly, not $L^\infty$).

Now for some remarks and results. We will not prove any of these here; some can be proven using techniques from Advanced Calculus, while others require Lebesgue theory. But all these results are good to know, and many are important for proofs used in this class.

- If a function $f$ is bounded and also has bounded support on $\mathbb{R}^n$, then $f$ is $L^\infty$ and $L^p$ for all $p$.
- The set $C_0(\mathbb{R}^n)$, and, hence, also $C_0^k(\mathbb{R}^n)$ for any $k$, meet the above requirements, and are, therefore, $L^\infty$ and $L^p$ for all $p$.

- The Schwartz space $S(\mathbb{R}^n)$ is a dense subset of $L^\infty(\mathbb{R}^n)$ and $L^p(\mathbb{R}^n)$ for all $p$.
- For any function $f$ in $L^1(\mathbb{R})$ (or $L^p(\mathbb{R})$ for $1 \le p \le \infty$)), there is a sequence of 'step functions' $\{g_k\}$ with
$$g_k(x) = \sum_{j=1}^{N} c_j \chi_{[a_j,b_j)}(x)$$
such that $\{g_k\}$ converges to $f$ in $L^1(\mathbb{R})$ (or $L^p(\mathbb{R})$). A similar result holds on $\mathbb{R}^n$, except that the intervals are replaced by 'boxes,' or Cartesian products of intervals.
- The larger $p$, the more strict the local integrability requirements on $L^p(\mathbb{R}^n)$, and the less strict the decay restrictions as $x \to \infty$. Consider rational functions to see this. For example $f(x) = \frac{1}{x}$ is not in $L^1(\mathbb{R})$, but it is in $L^2(\mathbb{R})$.
- A consequence of this is that for $1 \le q \le p \le r$, every $L^q$ function which is also $L^r$ is also $L^p$.
- If we look only at some bounded domain $D \subset \mathbb{R}^n$ (technically, if the measure of $D$ is finite), then
$$L^\infty(D) \subset .... \subset L^{k+1}(D) \subset L^k(D) \subset ..... \subset L^1(D).$$

## 3. Modes of convergence of functions

Now that we have defined various norms for spaces of functions, we can discuss some of the different things that it can mean for a sequence of functions to converge. One key point we want to emphasize is that there are different notions of convergence—pointwise convergence, uniform convergence, and convergence under various norms—and when you either read or state yourself that a sequence of functions converges, you should always ask *in what sense* it converges.

### 3.1. Pointwise Convergence.

The simplest notion of convergence of a sequence of functions is that of pointwise convergence.

**Definition 3.1.** Let $D$ be any set and $A$ a metric space, and let $\{f_n\}$ be a sequence of functions $D \to A$. If there is a function $f : D \to A$ such that for each $x$ in $D$, the sequence $f_n(x)$ converges to $f(x)$ in A, then we say $\{f_n\}$ *converges pointwise* to $f$, and we call $f$ the *pointwise limit* of $\{f_n\}$.

When we talk about pointwise convergence of a sequence of functions, we do *not* mean that the functions converge in some normed function space, as will be the case when we discuss other kinds of convergence. When we discuss pointwise convergence, the only metric space of interest is the co-domain $A$, and all we mean is that for each given input $x \in D$, the output $f_n(x)$ eventually gets 'close' to $f(x)$ in A. Thus, we need not specify any function space before talking about pointwise convergence.

In this course, the domain will usually be $R^n$, and the co-domain will generally be the normed space $\mathbb{R}$ or $\mathbb{C}$. Before we move on to some exercises, lets look at a simple example of a pointwise converging sequence of functions.

**Example 3.2.** Let $f_n : \mathbb{R} \to \mathbb{C}$ be given by $f_n(x) = \frac{x}{n} + i\frac{x}{n}$. Show that $\{f_n\}$ converges pointwise to $f : \mathbb{R} \to \mathbb{C}$ given by $f(x) = 0$.
*Proof:* Choose $x \in \mathbb{R}$, and let $\epsilon > 0$. Choose $N \in \mathbb{N}$ such that $N > 2\frac{|x|}{\epsilon}$. Then for each $n \ge N$,
$$\begin{aligned} |f_n(x) - f(x)| &= \left|\frac{x}{n} + \frac{ix}{n}\right| \\ &\le \left|\frac{x}{n}\right| + \left|\frac{ix}{n}\right| \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2} \\ &= \epsilon. \end{aligned}$$

Note that, as usual when discussing convergence of sequences, our choice of $N$ will typically depend on $\epsilon$, that is, the smaller $\epsilon$, the larger $N$ will generally need to be.

**Example 3.3.** Find the pointwise limits, if they exist, of the following sequences of functions. Be careful of the endpoints of intervals.

(a) $f_n : [0, \infty) \to \mathbb{R}$, where $f_n(x) = x^{1/n}$,
(b) $f_n : [0, 1] \to \mathbb{R}$, where $f_n(x) = \sin(nx)$,
(c) $f_n : \mathbb{R} \to \mathbb{R}$, where $f_n(x) = \sin \frac{x}{n}$,
(d) $f_n : (0, 1] \to \mathbb{R}$, where
$$f_n(x) = \begin{cases} n, & 0 < x \le \frac{1}{n}, \\ 0, & \text{otherwise.} \end{cases}$$
(e) $f_n : [0, 1] \to \mathbb{R}$, where $f_n(x) = x^n$,

(f) $f_n : [0,1] \to \mathbb{R}$, where

$$f_n(x) = \begin{cases} 1 & \text{if there exist } \quad p, q \in \mathbb{Z} \text{ with } q \leq n \text{ and } x = \frac{p}{q}, \\ 0, & \text{otherwise.} \end{cases}$$

*Solution:*

(a) For each $n$, $f_n(0) = 0$, and for any $x > 0$, $f_n(x) \to 1$.
(b) The function $f_n$ oscillates tighter and tighter as $n \to \infty$, so the pointwise limit does not exist.
(c) For each $x \in \mathbb{R}$, $\lim\limits_{n\to\infty} x/n = 0$, so by continuity of sin, for each $x \in \mathbb{R}$ we have $f_n(x) \to \sin(0) = 0$.
(d) For each $x > 0$, there is $N \in \mathbb{N}$ such that whenever $n \geq N$, $\frac{1}{n} < x$. It follows that $\lim\limits_{n\to\infty} f_n(x) = 0$ for each $x \in (0,1]$. Note that if we used a closed interval, we would either have to say there is no pointwise limit or allow the limit of $f_n(0)$ to be $\infty$.
(e) For every $n \in \mathbb{N}$, $f_n(1) = 1$, and for each $x \in [0,1)$ we have $\lim\limits_{n\to\infty} f_n(x) = 0$.
(f) For every $r \in \mathbb{Q}$, there is $N \in \mathbb{N}$ sufficiently large that whenever $n \geq N$, $f_n(r) = 1$, so we have $f_n(r) \to 1$. For any irrational number $x$ there exists no such $n$, so we have $f_n(x) \to 0$. Hence, this sequence of functions converges to the Dirichlet function, which is 1 on the rationals and 0 on the irrationals.

□

*Remark* 3.4. The above examples demonstrate some problems with pointwise convergence. We would like to be able to determine properties of the limit of a sequence of functions by looking only at the terms of the sequence itself. For example, we would like to be able to say that the limit of a sequence of continuous functions is continuous, that the limit of a sequence of Riemann integrable functions is Riemann integrable, and that the integral of the limit of a sequence of functions is the limit of their integrals. But the above exercise contains counterexamples showing all three of these statements to be false.

**Example 3.5.** We can justify Remark 3.4 using functions from Example 3.3. In particular,

- Both (a) and (e) above suffice to show that the pointwise limit of a sequence of continuous functions need not be continuous: in either case, $f_n$ is continuous on its entire domain, but the limit function $f$ has a discontinuity. In (a), the discontinuity is at 0, while in (e) it is at 1.
- The sequence in (f) shows that the pointwise limit of a sequence of Riemann integrable functions need not be Riemann integrable: for all $n$, $f_n$ has only finitely many jump discontinuities, and has value of zero at all but these finitely many points, so $f_n$ is Riemann integrable with integral 0. But the limit function is discontinuous everywhere, so no matter how small a partition we take we will always get an upper sum of one and lower sum of zero, which implies that the limit function is not Riemann integrable.
- The sequence in (d) shows that, even if the pointwise limit $f$ of a sequence of Riemann integrable functions $f_n$ is Riemann integrable, it need not be true that

$$\lim_{n\to\infty} \int_a^b f_n(x)dx = \int_a^b f(x)dx.$$

Since the limit function is identically zero, it must be integrable with integral zero. But calculating the area under the 'box' at the far left side of the domain of each $f_n$ reveals that for every $n$,

$$\int_0^1 f_n(x)dx = n\frac{1}{n} = 1.$$

## 3.2. Uniform Convergence.

**Definition 3.6.** Let $\{f_n\}$ be a sequence of functions from a set $D$ to a metric space $A$. We say that $\{f_n\}$ *converges uniformly* to $f : D \to A$ if for each $\epsilon > 0$ there is an $N \in \mathbb{N}$ such that for every $n \geq N$ and for every $x \in D$,

$$d(f(x), f_n(x)) < \epsilon.$$

In this case, we call $f$ the *uniform limit* of $\{f_n\}$.

First, note that uniform convergence is, in fact, a stronger condition than pointwise convergence. That is, we can see from the definition that any sequence of functions $\{f_n\}$ which converges converges uniformly to a function $f$, will also converges pointwise to $f$. The converse is not true. For example the function

$$\chi_{[0,\frac{1}{n}]}$$

converges pointwise, but not uniformly, to 0, because for any $n$ there is some $x$ such that $|f_n(x) - 0| = 1$.

The key difference between uniform and pointwise convergence is that if we have uniform convergence, we can pick $N$ based solely on knowing $\epsilon$, without regard to $x$, that is, the same value of $N$ suffices for every $x$. Loosely speaking, uniform continuity means that we can squeeze the tail end of the sequence of functions into a tube of radius $\epsilon$ about the limit function $f$ for arbitrarily small $\epsilon$. In the case of functions $\mathbb{R} \to \mathbb{R}$, this is particularly easy to visualize, since in the case of the real line, this just means that for all $x \in \mathbb{R}$, $f(x) - \epsilon < f_n(x) < f(x) + \epsilon$.

Uniform convergence preserves a number of nice properties, including continuity and Riemann integrability (on finite intervals; not necessarily improper Riemann integrability). We leave the proofs of both facts to the reader.

As with pointwise convergence, we can talk about uniform convergence almost everywhere if the domain is a measure space. In this case, the sequence converges uniformly except on a set of measure zero.

### 3.3. Convergence in Normed Function Spaces.

Recall that sequence $\{x_n\}$ in a normed space $(X, \|\cdot\|)$ converges to $x \in X$ if for every $\epsilon > 0$ there is an $N \in \mathbb{N}$ such that for each $n \geq \mathbb{N}$, $\|x - x_n\| < \epsilon$.

We now apply this definition to sequences of functions in normed spaces. The first example we already did in the last section: if we restrict our attention to bounded functions, uniform convergence can be identified with convergence in the space of bounded functions under the uniform norm. Similarly, almost-everywhere uniform convergence is equivalent to convergence under the $\infty$ norm. Pointwise convergence, on the other hand, is not equivalent to convergence under any metric. It is an example of the more general concept of convergence in a topological space, which we will not define here.

Uniform convergence is a fairly straightforward concept. Among other things, uniform convergence implies pointwise convergence. $L^p$ convergence, on the other hand, is more subtle. In order to converge in the $L^1$ sense, for example, a sequence of functions need not converge uniformly, or even pointwise. In fact, as we shall see in one of the following exercises, a function can converge in the $L^1$ without converging at *any* point in its domain. At the same time, it is possible for a sequence to converge pointwise, or even uniformly, without converging in the $L^1$ sense.

**Example 3.7.** Do the following:

(a) Find a sequence of functions which converges uniformly on an unbounded domain but fails to converge in the $L^1$ sense.

(b) Find a sequence of functions which converges in the $L^1$ sense, but does not converge pointwise *anywhere* in the domain.

*Solution:*

(a) Let $f_n : \mathbb{R} \to \mathbb{R}$ be defined by $f_n(x) = \frac{1}{n}$. Let $f(x) = 0$. Then $f_n \to f$ uniformly, but not in the $L^1$ sense.

(b) Define $f_n : [0,1] \to \mathbb{R}$ by

$$f_n(x) = \begin{cases} 1, & x \in A_n, \\ 0, & \text{otherwise}, \end{cases}$$

where $A_1 = [0,1]$, $A_2 = [0, \frac{1}{2}]$, $A_3 = [\frac{1}{2}, 1]$, $A_4 = [0, \frac{1}{3}]$, $A_5 = [\frac{1}{3}, \frac{2}{3}]$, and so on. It is clear that, since $f_n = \chi_{A_n}$ and the length of the $A_n$'s approaches zero, we have $\int_0^1 (f_n - 0)dx \to 0$, so $f_n \to 0$ in $L^1$. But for any $x \in [0,1]$, we can see that $f_n(x) = 1$ for infinitely many values of $n$, so the sequence $\{f_n\}$ does not converge pointwise *anywhere* in [0,1].

□

Nonetheless, the Lebesgue Dominated Convergence Theorem does show that there is a connection between pointwise convergence and $L^p$ convergence; this was one of the primary motivations for the development of the Lebesgue integral.

## 4. An Example of Topological Convergence: The Schwarz Class

Not all forms of convergence can be described in terms of norms. One example is pointwise convergence. Another is convergence in the Schwarz class, which we describe in this section. These other forms of convergence are called *topological convergence*. A topology is an analytic tool in which open sets, rather than a metric, are taken as the fundamental unit. We will not worry about details from general topology here, but describe the specific case of Schwarz functions, which will play a major role later when we discuss convolutions, distributions, and the Fourier transform.

Earlier, we stated that the Schwarz class of functions on $\mathbb{R}$ is the subset of functions in $C^\infty(\mathbb{R})$ that are rapidly decreasing and have all derivatives rapidly decreasing. In this section, we introduce the class $\mathcal{S}(\mathbb{R}^n)$, the set of rapidly decreasing functions on $\mathbb{R}^n$ with rapidly decreasing derivatives of all orders. To make this precise, we introduce some notation that will be of use when we study the Fourier transform.

A *multi-index* is an ordered $n$-tuple of nonnegative integers. For a multi-index $\alpha = (\alpha_1, \alpha_2, ..., \alpha_n)$, we define

$$|\alpha| = \sum_{j=1}^n \alpha_j, \quad \alpha! = \prod_{j=1}^n \alpha_j!, \quad \partial^\alpha = \left(\frac{\partial}{\partial x_1}\right)^{\alpha_1} \cdots \left(\frac{\partial}{\partial x_n}\right)^{\alpha_n},$$

and for $x \in \mathbb{R}^n$,

$$x^\alpha = \prod_{j=1}^n x_j^{\alpha_j}.$$

With this notation, for example, the product rule for derivatives can be written

$$\partial^\alpha(fg) = \sum_{\beta+\gamma=\alpha} \frac{\alpha!}{\beta!\gamma!}(\partial^\beta f)(\partial^\gamma g).$$

When we are discussing a first-order partial derivative, we will find it more convenient to use the notation $\partial_j$, where $j$ is the component with which we are taking a partial derivative.

In this section and in our study of the Fourier transform, we will use the absolute value $|x|$ to denote the Euclidean norm of a vector in $\mathbb{R}^n$, in order to avoid confusing norms on vectors with norms on functions. The notation is inconsistent with the notation we use on multi-indexes, but it will be clear from context whether we are considering a particular $n$-tuple to be a real vector or a multi-index.

We now define semi-norms

$$\|f\|_{N,\alpha} = \sup_{x\in\mathbb{R}^n} (1+|x|)^N |\partial^\alpha f(x)|,$$

where $N \in \mathbb{N}$ and $\alpha$ is a multi-index, and we define the Schwarz class to be

$$\mathcal{S}(\mathbb{R}^n) = \{f \in C^\infty(\mathbb{R}^n) : \|f\|_{N,\alpha} < \infty \quad \text{for all} \quad N, \alpha\}.$$

These functions and their derivatives decay faster than any rational function.

**Example 4.1.** The standard example of a Schwarz function is the Gaussian, $e^{-|x|}$. Any multiple of the Gaussian and a polynomial is also a Schwarz function, as are any functions in $C_c^\infty(\mathbb{R}^n)$, such as the 'bump' function on $\mathbb{R}$ defined by

$$f(x) = \begin{cases} e^{-1/(1-|x|)^2} & -1 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

**Proposition 4.2.** *Suppose $f \in C^\infty$. Then $f \in \mathcal{S}(\mathbb{R}^n)$ if and only if $x^\beta \partial^\alpha f(x)$ is bounded for all multi-indexes $\alpha$ and $\beta$ if and only if $\partial^\alpha(x^\beta f(x))$ is bounded for all multi-indexes $\alpha, \beta$.*

*Proof.* Since $\left|x^\beta\right| \le (1+|x|)^N$ for $|\beta| \le N$, the first 'only if' is easy. Conversely, since $\sum_{j=1}^n |x_j|^N$ is strictly positive on the unit sphere, so by compactness[2] it has a strictly positive minimum $\delta$ there. Thus, $\sum_{j=1}^n |x_j|^N \ge \delta |x|^N$ for all $x$, because both sides are homogeneous of degree $N$[3], and so we find that

$$(1+|x|^N) \le 2^N(1+|x|^N) \le 2^N\left(1+\frac{1}{\delta}\sum_{j=1}^n |x_j|^N\right) \le 2^N \frac{1}{\delta}\sum_{|\beta|\le N}\left|x^\beta\right|,$$

proving the first equivalence. The second follows from the fact that, by the product rule and the fact that the derivative of a polynomial is another polynomial, any term $\partial^\alpha(x^\beta f)$ is a linear combination of terms $x^\gamma \partial^\eta f$ and vice-versa. $\square$

Now we reach the reason for putting our discussion of the Schwarz space in this section on convergence in function spaces: we define what it means for Schwarz functions to converge. Like pointwise convergence, convergence in the Schwarz space is a type of topological convergence which cannot be described by a metric. As one might suspect from our definition of the space, a sequence of functions $f_n \in \mathcal{S}(\mathbb{R}^n)$ is said to converge to a function $f \in \mathcal{S}(\mathbb{R}^n)$ if, for each $N \in \mathbb{N}$ and for each multi-index $\alpha$, we have

$$\|f_n - f\|_{N,\alpha} \to 0.$$

A linear function $l$ from $\mathcal{S}(\mathbb{R}^n)$ to a metric space $X$ is said to be *continuous* if for any $f \in \mathcal{S}(\mathbb{R}^n)$ and $f_n \in \mathcal{S}(\mathbb{R}^n)$ such that $f_n \to f$, we have $l(f_n) \to l(f)$. Such linear functions are also called *bounded*, although we have to be careful to note that the word bounded, when applied to linear functions, means something different than when applied to general functions, as no nonzero linear function can ever be bounded in the usual sense. The space of such linear functions which are continuous on $\mathcal{S}(\mathbb{R}^n)$ is denoted $\mathcal{S}'(\mathbb{R}^n)$, and as we will see plays a role in the theory of convolutions and the Fourier transform.

[2]This is the $n$-dimensional version of the theorem from Advanced Calculus stating that the image of a compact domain in a continuous function is compact

[3]A function is homogeneous of degree $N$ if changing the scale of the inputs by a factor $\gamma$ has no other effect than to rescale the output by a factor of $\gamma^N$.

## 5. Weak derivatives

One last topic we should cover in our survey of $L^p$ theory is the idea of a weak derivative, often called an $L^p$ derivative when discussing $L^p$ functions. The idea is to use integration by parts, a consequence of the theory of classical derivatives, to *define* a derivative by demanding only that we can integrate by parts. There are two advantages to using these derivatives on $L^p$:

- Like the $\|\cdot\|_\infty$ norm in contrast to the uniform norm, a derivative of $f$ defined in terms of integration will not be sensitive to pointwise behavior of $f$. Given that the functions in $L^p$ are technically equivalence classes defined only modulo sets of measure zero, this is rather important.
- Such a derivative will be slightly more general than the classical derivatives studied in advanced calculus. For example, a continuous function with a classical derivative defined at all but finitely many points will have a weak derivative equal to that classical derivative, since we can integrate by parts everywhere that the derivative exists, and the rest of the domain will be a set of measure zero. Hence, functions like $|\cdot|$ will be have weak derivatives.

Before we define a weak derivative, we review multi-index notation, often the most convenient way to denote partial derivatives.

**Definition 5.1.** Let $\alpha \in \mathbb{R}^n$ with each $\alpha_j$ a nonnegative integer. We call $\alpha$ a *mutli-index*, and write $|\alpha|$ for $\sum_{j=1}^n |\alpha_j|$. If $x \in \mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}$, we define

$$x^\alpha = \prod_{j=1}^n x_j^{\alpha_j}$$

and

$$\partial^\alpha f = \left(\frac{\partial_1^\alpha}{\partial x_1^{\alpha_1}}\right)\left(\frac{\partial_2^\alpha}{\partial x_2^{\alpha_2}}\right)\cdots\left(\frac{\partial_n^\alpha}{\partial x_n^{\alpha_n}}\right) f.$$

**Definition 5.2.** Let $f$ and $g$ be locally integrable functions on $U \subset \mathbb{R}^n$ and $\alpha$ a multi-index. We call $g$ the $\alpha$'th *weak (partial) derivative* of $f$ on $U$ if, for $\phi \in C_c^\infty(U)$, that is, for every infinitely differentiable $\phi$ with compact support in $U$, we have

$$\int_U f D^\alpha \phi = -(1)^{|\alpha|} \int_U g\phi.$$

In Fourier analysis and distribution theory, it is usually the idea of a weak derivative which we use, since the results generally involve integrating one function against another.

## 6. Convolutions of functions on $\mathbb{R}^n$

The presentation in this section is adapted from [2] and [1].

We are now almost prepared to study the Fourier transform for functions defined on $\mathbb{R}^n$. Before we do, however, it will be convenient to already have some results regarding convolutions of functions on $\mathbb{R}^n$. As with convolutions over $\mathbb{Z}_n$, a convolution over $\mathbb{R}^n$ can be thought of as a weighted average, this time in an integral. On the other hand, it is not true that all shift-invariant linear transformations on functions of $\mathbb{R}^n$ can be represented by a convolution with a function; not even the identity, carrying a function $f$ to itself, can be expressed as a convolution without using generalized functions, which we will describe after we look at the basic properties of convolutions.

**Definition 6.1.** Let $f$ and $g$ be measurable functions. The *convolution* of $f$ and $g$ is the function $f * g$ defined for $x \in \mathbb{R}^n$ by

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

**Definition 6.2.** If $k \in \mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{C}$, define the *translate* of $f$ by $k$, $R_k f$, by

$$R_k f(x) = f(x-k).$$

The convolution is clearly linear in both arguments. Its other basic properties are summarized below.

**Proposition 6.3.** *Under the assumption that all the integrals converge absolutely, the convolution satisfies the following properties:*

i. $f * g = g * f$.
ii. $(f * g) * h = f * (g * h)$.
iii. *For any* $k \in \mathbb{R}^n, R_k(f * g) = (R_k f) * g = f * (R_k g)$.
iv. *If $A$ is the closure of the set* $\{x + y : x \in \operatorname{supp}(f), y \in \operatorname{supp}(g)\}$, *then* $\operatorname{supp}(f * g) \subset A$.

*Proof.* (i) We make the substitution $z = x - y$, to get

$$f * g(x) = \int_{-\infty}^{\infty} f(x-y)g(x)dy = \int_{-\infty}^{\infty} f(z)g(x-z)dz = g * f(x).$$

(ii) We use (i) and Fubini's theorem to get

$$\begin{aligned}(f*g)*h(x) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(y)g(x-z-y)h(z)dydz \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(y)f(x-y-z)h(z)dzdy \\ &= f*(g*h)(x).\end{aligned}$$

(ii) The first equality follows directly from the definitions:

$$R_k(f*g)(x) = f*g(x-k) = \int_{-\infty}^{\infty} f(x-k-y)g(y)dy = \int_{-\infty}^{\infty} R_k f(x-y)g(y)dy.$$

The second follows from the first and (i).

(iii) This is true because if $x \notin A$, then for any $y \in \text{supp}(f), x - y \notin \text{supp}(g)$, so that either $f(x-y)$ or $g(y)$ will be zero for every $y \in \mathbb{R}^n$. Thus, the integral defining $f * g(x)$ must be zero. □

Now we must address the question of when these integrals do exist. We will present these results without proof, as they require some technical details to prove

**Proposition 6.4.** Young's Inequality *If $f \in L^1$ and $g \in L^p$ for $1 \le p \le \infty$, then the integral defining $f * g(x)$ is absolutely convergent for almost every $x$, $f * g \in L^p$, and $\|f*g\|_p \le \|f\|_1 \|g\|_p$.*

**Proposition 6.5.** *If $\frac{1}{p} + \frac{1}{q} = 1$, and $f \in L^p, q \in L^q$, then the integral defining $f * g(x)$ is absolutely convergent for every $x$. Moreover, $f * g$ is bounded and uniformly continuous, with $\|f*g\|_u \le \|f\|_p \|g\|_q$. If $p$ and $q$ are strictly between 1 and infinity, then also $f * g \in C_0(\mathbb{R})$.* [4]

One property of convolutions which makes them so useful is their *smoothing* properties; the convolution product generally satisfies the same regularity conditions as the more regular of the two functions in the product

**Proposition 6.6.** *If $f \in L^1$ and $g$ is continuous and bounded, then $f * g$ is continuous. Similarly, if $g$ is $C^k$ and $\partial^\alpha g$ is bounded for each $|\alpha| < k$, then $f * g$ is $C^k$, and partial$^\alpha(f*g) = f * (\partial^\alpha g)$.*

*Proof.* We present the argument for the derivatives; the argument for continuity is similar, but simpler because there is no difference quotient. In both cases, the boundedness criterion is used in order to apply the Dominated Convergence Theorem

Since we can proceed inductively, it suffices to prove the theorem for $\partial_j f * g$. For brevity, denote $\tau = te_j$ in the argument below:

$$\lim_{t\to 0} \frac{f*g(x+\tau) - f*g(x)}{t} = \lim_{t\to 0} \int_{-\infty}^{\infty} \left( \frac{g(x+\tau-y) - g(x-y)}{t} \right) f(y).$$

We can then use the Dominated Convergence Theorem to bring the limits inside the integral, since by the mean value theorem all of the difference quotients are bounded by $\sup_{x\in\mathbb{R}^n} |\partial_j g(x)|$, and $\sup_{x\in\mathbb{R}^n} |\partial_j g(x)| \, |f(y)|$ is $L^1$. □

We already know from Young's inequality that $L^1(\mathbb{R}^n)$ is closed under convolutions. We now show that $\mathcal{S}(\mathbb{R}^n)$ is, as well.

**Proposition 6.7.** *If $f, g \in \mathcal{S}(\mathbb{R}^n)$, then $f * g \in \mathcal{S}(\mathbb{R}^n)$.*

*Proof.* By Proposition 6.5, $f * g$ is clearly defined, and by Proposition 6.6 it is $C^\infty$, so we need only show that $\|f*g\|_{N,\alpha}$ is finite for all $N, \alpha$. We note that since, by the triangle inequality

$$1 + |x| \le 1 + |x-y| + |y| \le (1+|x-y|)(1+|y|),$$

[4] Recall that $C_0$ is the set of continuous functions that go to zero as $|x| \to \infty$.

we have

$$\begin{aligned}
&(1+|x|)^N|\partial^\alpha(f*g)(x)| \\
&\qquad \leq \int_{-\infty}^{\infty}(1+|x-y|)^N|\partial^\alpha f(x+y)|(1+|y|)^N\,|g(y)|\,dy \\
&\qquad \leq \int_{-\infty}^{\infty}\|f\|_{N,\alpha}\,\|g\|_{N+n+1,0}\,(1+|y|)^{-n-1}dy \\
&\qquad \leq \|f\|_{N,\alpha}\,\|g\|_{N+n+1,0}\int_{-\infty}^{\infty}(1+|y|)^{-n-1}dy.
\end{aligned}$$

The integral above is finite (this is reasonably straightforward using polar coordinates) and does not depend on $x$, so, by taking the supremum over $x$, we get the desired result. □

It should be clear to the reader that there is no exact identity in convolutions; any such function $g$ would have to satisfy the condition $\int f(x-y)g(y)dy = f(x)$ for all $f$, which is not possible, since this would require that $g$ have unit mass at 0, and the integral is defined such that this cannot occur. On the other hand, this condition should give us the idea that, by squeezing more and more mass near 0, we can approximate a convolution identity, at least for continuous functions

**Proposition 6.8.** *Let $g \in L^1$ with $\|g\|_1 = 1$, and define $g_t$ by $g_t(x) = \frac{1}{t}g(x/t)$. Then, as $t \to 0$,*

i. *For $f \in L^p$, $1 \leq p < \infty$, $f * g_t \to f$ in the $L^p$ norm.*
ii. *For any $f$ which is bounded and uniformly continuous, $f * g_t \to f$ uniformly.*
iii. *For $f \in L^\infty$, continuous on an open set $U$, $f * g_t \to f$ uniformly on compact subsets of $U$.*

*Proof.* The proof of the first fact is a bit technical, but we will prove the second fact, which implies the third because continuity on a compact set implies uniform continuity and boundedness on that set.

Let $f$ be bounded by $M$ and uniformly continuous. Let $\epsilon > 0$. Note that for $t \neq 0$, $\|g_t\|_1 = 1$. Then

$$\begin{aligned}
|f * g_t(x) - f(x)| &= \left|\int_{-\infty}^{\infty}(f(x-y)-f(x))g_t(y))dy\right| \\
&\leq \int_{-\infty}^{\infty}|f(x-tz)-f(x)|\,|g(z)|dz \\
&= \int_{|z|\leq N}|f(x-tz)-f(x)|\,|g(z)|dz + \int_{|z|>N}|f(x-tz)-f(x)|\,|g(z)|dz \\
&\leq \sup_{|w|\leq N}(|f(w-tN)-f(w)|\,\|g\|_1 + 2\,\|f\|_u\int_{|z|>N}|g(z)|dz,
\end{aligned}$$

where we made a change of variables in the second line, partitioned the domain of integration in the the third (this is technically justified in the case of Lebesgue integration by linearity, since the integral of a function $h$ over $A \subset \mathbb{R}^n$ is defined to be $\int_{\mathbb{R}^n} f\chi_H$) and used monotonicity in the last.

The second term converges to 0 as $N \to \infty$, and for any given $N$ the first term converges to 0 as $t \to 0$ by uniform continuity. Neither term's convergence depends on the value of $x$, so the convergence is uniform. □

Proposition 6.8 gives us some idea of how to approximate an identity in convolutions, but we would like to have a precise identity, that is, some function, or rather some 'generalized function,' with the property that it has integral one precisely at the point zero. This will require a new theory of functions, first introduced in the 1930's, called distributions.

## 7. Distributions and the Convolution Identity

Distributions, or generalized functions, are defined in terms of how they integrate against other functions. The general idea is as follows: given a space of integrable functions $S$, we associate with it $S'$, the space of continuous linear maps from $S$ to $\mathbb{C}$. Such linear maps are called linear *functionals.* Heuristically, we imagine the value of a functional $f$ at a Schwarz function $\phi$ to be the integral $\int f\phi$. This integral will define a continuous linear functional whenever $f$ is of a suitable class of locally integrable functions, but for many distributions $f$ there will be no ordinary function $g$ satisfying $f(\phi) = \int g\phi$, which is why we call $f$ a generalized function.

Distributions are used for a number of purposes, but three of the main reasons are that they extend the definition of derivatives, they extend the definition of convolutions, and they extend the definition of the Fourier transform to a wide variety of functions. We will outline how the first two are defined below, and cover the third after we cover the Fourier transform for ordinary functions.

We will work with tempered distributions, defined below:

**Definition 7.1.** A linear functional $f : \mathcal{S}(\mathbb{R}^n) \to \mathbb{C}$ is called a *tempered distribution* if it is continuous on $\mathcal{S}(\mathbb{R}^n)$. Continuity here means that for any sequence of Schwarz functions $\phi_n$ converging to a Schwarz function $\phi$ in the topology on $\mathcal{S}(\mathbb{R}^n)$ (i.e., converging in all of the semi-norms $\|\cdot\|_{N,\alpha}$) we have $f(\phi_n) \to f(\phi)$ in $\mathbb{C}$. The space of tempered distributions on $R^n$ is denoted $\mathcal{S}'(\mathbb{R}^n)$, and by convention the output $f(\phi)$ is instead written in inner product notation as $\langle f, \phi \rangle$.

We equip $\mathcal{S}'(\mathbb{R}^n)$ with the weak* topology. Without worrying about details of topology, this means that a sequence of tempered distributions $f_n$ converges to $f$ in $\mathcal{S}'(\mathbb{R}^n)$ if it converges pointwise on $\mathcal{S}(\mathbb{R}^n)$, that is, if $\langle f_n, \phi\rangle \to \langle f, \phi\rangle$ in $\mathbb{C}$ for every $\phi \in \mathcal{S}(\mathbb{R}^n)$.

*Remark* 7.2. In the interests of saving time, we will omit the proofs of continuity for the distributions we define below. Most of the details can be found in [2].

**Example 7.3.** The example which motivates our use of tempered distributions as generalized functions is that of any function on $\mathbb{R}^n$ which is locally integrable and has at most polynomial growth, which we will call a *tempered* function. Note that this includes all the $L^p$ spaces, as well as all polynomials and trigonometric polynomials. We can associate any such function $f$ with a tempered distribution, which we will also call $f$, setting

$$\langle f, \phi\rangle = \int_{-\infty}^{\infty} f(x)\phi(x)dx.$$

Also, note that if we define $\tilde{\phi}(x) = \phi(-x,)$, then we also have

$$\langle f, \phi\rangle = f * \tilde{\phi}(0),$$

which suggests that we can use distributions to extend the definition of a convolution. We equip $\mathcal{S}'(\mathbb{R}^n)$ with the topology of pointwise convergence. That is, we say that a sequence of tempered distributions $f_n$ converges to a tempered distribution $f$ if, for every $\phi \in \mathcal{S}(\mathbb{R}^n)$,

$$f_n(\phi) \to f(\phi).$$

**Example 7.4.** One of the most important tempered distributions is the Dirac delta distribution, or point mass, defined by $\langle \delta, \phi\rangle = \phi(0)$.

**Corollary 7.5.** *For any sequence $g_t$ as in Proposition 6.8 we have that $g_t$, viewed as a sequence of elements of $\mathcal{S}'(\mathbb{R}^n)$, converges to $\delta$.*

*Proof.* Details are left to the reader. □

**Example 7.6.** For any multi-index $\alpha$, the operator $\partial^\alpha$ defines a tempered distribution.

Distribution theory is usually developed first on the space $\mathcal{D}(U) = C_c^\infty(U)$, the continuous functions with compact support in some open set $U$. This is done for several reasons, in particular because the space $C_c^\infty$ is particularly easier to work with. However, since our ultimate goal is to be able to perform Fourier transforms on generalized functions, and this is not possible for every element of $\mathcal{D}'(\mathbb{R}^n)$, we jump immediately to the tempered distributions. Of course, every tempered distribution is also a distribution, since the domain of tempered distributions includes the domain of distributions, but there are some distributions which are not tempered. For example, any locally integrable function $f$ can be associated with a continuous linear functional on $\mathcal{D}(\mathbb{R}^n)$ by mapping $\phi$ to

$$\int_{-\infty}^{\infty} f(x)\phi(x)dx,$$

since the integral is certain to converge, but some (not all) functions that grow faster than polynomials cannot be associated with elements of $\mathcal{S}'(\mathbb{R}^n)$ in that manner. For example, $e^{x^2}$ is locally integrable on $R^1$, but

$$\int_{\mathbb{R}^1} e^x e^{-x^2/2} dx$$

does not converge.

Before we begin defining basic operations for distributions, we remark that, like the *delta* function, all distributions can be approximated by ordinary functions. In fact, viewed as a subset of $\mathcal{S}'(\mathbb{R}^n)$ using the integration formula given above, $\mathcal{S}(\mathbb{R}^n)$ itself is dense in $\mathcal{S}'(\mathbb{R}^n)$, and in fact even $C_c^\infty$ is.

7.1. **Operations on Distributions.** Below, we outline how to extend a number of basic properties of functions to distributions. Although we will omit proofs that these operations produce linear functionals which are continuous on $\mathcal{S}(\mathbb{R}^n)$, we will check that each of these definitions is consistent with the already existing definitions for functions in the Schwarz space. To be precise, we use the following procedure:

Suppose we have a linear operation $T$ we would like to be able to perform on a distribution $f$. We look for another linear operation $T'$ such that, if $f \in \mathcal{S}(\mathbb{R}^n)$, then for any $\phi \in \mathcal{S}(\mathbb{R}^n)$ we have

$$\int (Tf)g = \int f(T'g).$$

We can then *define* $Tf$, for general tempered distributions $f$, by

$$\langle Tf, \phi \rangle = \langle f, T'\phi \rangle ,$$

assured that the definition will be consistent with the existing definition when $f$ is also a function.

We demonstrate this process for several basic operations:

- Let $Tf = \partial^\alpha f$, defined on the subset of $C^{|\alpha|}(\mathbb{R})$ such that all partial derivatives of orders less than or equal to $\alpha$ are tempered functions. Then, since $\phi$ decreases rapidly to 0, Fubini's theorem and integration by parts repeated $|\alpha|$ times gives $\int (\partial^\alpha f)\phi = (-1)^{|\alpha|} \int f(\partial^\alpha \phi)$, since the boundary terms vanish. Noting that $\partial^\alpha \phi$ is also Schwarz, we then define the operator $\partial^\alpha$ for $f \in \mathcal{S}'(\mathbb{R}^n)$ by
$$\langle \partial^\alpha f, \phi \rangle = (-1)^{|\alpha|} \langle f, \partial^\alpha \phi \rangle .$$
Note that with this definition, all tempered distributions are infinitely differentiable. Also note that, by treating tempered functions—a wide class that, as we mentioned before, contains all the $L^p$ functions and many others—as tempered distributions, then we can define derivatives of all orders for any such functions, although the derivatives will in general be generalized functions rather than functions. In the cases where a tempered function $f$ has a classical derivative or even a weak derivative $g$, then the distributional derivative will be equivalent to the weak derivative.
- Suppose that $\psi$ is a $C^\infty$ function for which all derivatives are bounded, and let $Tf = \psi f$, the product of $f$ and $\psi$. If $f$ is a tempered function and $\phi \in \mathcal{S}(\mathbb{R}^n)$, then $\int \psi f \phi = \int f \psi \phi$. Noting that $\psi\phi$ is also Schwarz, we then define, for general tempered distributions $f$,
$$\langle \psi f, \phi \rangle = \langle f, \psi \phi \rangle .$$
- Recall that $R_k$ is the translation operator by $k$. Let $Tf = R_k f$. It follows from a simple change of variables that if $f$ is a tempered function and $\phi \in \mathcal{S}(\mathbb{R}^n)$, then $\int (R_k f)\phi = \int f R_{-k}\phi$, so for general tempered distributions $f$ we define
$$\langle R_k f, \phi \rangle = \langle f, R_{-k}\phi \rangle .$$
For example, $R_k\delta$ is the point mass at $k$.
- Let $S$ be an invertible linear map on $\mathbb{R}^n$, and let $Tf = f \circ S$. A simple application of the linear change of variables formula, which we leave to the reader, shows that $T'\phi = |\det S|^{-1} \phi \circ S^{-1}$, and hence for general $f \in \mathcal{S}'(\mathbb{R}^n)$ we define
$$\langle f \circ S, \phi \rangle = |\det S|^{-1} \langle f, \phi \circ S^{-1} \rangle .$$

*Remark* 7.7. One of the primary reasons for working with tempered distributions is that, with the above definition, *every tempered distribution is infinitely differentiable.* Since the tempered distributions include locally integrable tempered functions, this means that the vast majority of the functions which we work with now have derivatives, in a generalized sense.

It is worth noting that, given the technique we used to define distributional derivatives, this derivative generalizes not only the notion of a classical derivative, but also that of a weak derivative.

A few examples should help the idea of a distributional derivative:

**Example 7.8.** The derivative of the step function,

$$g(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases} ,$$

is the delta function $\delta$. By linearity and shift-invariance of the derivative, the derivative of any function that has a weak derivative except at finitely many jump discontinuities will be the sum of that weak derivative and finitely many translates of multiples of the dirac delta function.

The derivative of $\delta$ is the distribution $\delta'$ defined by

$$\langle \delta', \phi \rangle = \phi'(0).$$

Similarly, the derivative of $\delta'$ is the functional that outputs $\phi''(0)$.

We leave the proofs to the exercises.

A useful consequence of the smoothing effect of convolutions is that we can define the convolution of a distribution and a Schwarz function in such a way as to get a complex-valued function, rather than simply a tempered distribution as the method used above would give.

- Let $f$ be a tempered function and $\psi \in \mathcal{S}(\mathbb{R}^n)$. Then, using a change of variables, $f * \psi(x) = \int f(x-y)\psi(y)dy = \left\langle f, R_x\tilde{\psi} \right\rangle$. Translations of Schwarz functions are Schwarz, so we can define, for an arbitrary tempered distribution $f$, a complex-valued function $f * \psi$ by
$$f * \psi(x) = \left\langle f, R_x\tilde{\psi} \right\rangle.$$

We can now check explicitly that
$$\delta * \phi(x) = \left\langle \delta, R_x\tilde{\phi} \right\rangle = R_x\tilde{\phi}0 = \tilde{\phi}(-x) = \phi(x),$$
so that $\delta$ is the convolution identity.

Of course, a downside of the definition used above is that, although we know that the convolution is a function, we are not assured that it will be a tempered function and hence do not know for certain that it will give a tempered distribution. However, it can be shown that if $f \in \mathcal{S}'(\mathbb{R}^n)$ and $\psi \in \mathcal{S}(\mathbb{R}^n)$, then the function $f * \psi$ is an infinitely differentiable function, and all its derivatives have at most polynomial growth. Thus, $f * \psi$ and its classical derivatives all define tempered distributions.

## References

[1] Charles Epstein. *Introduction to the Mathematics of Medical Imaging*. Prentice Hall, Upper Saddle River, 2nd edition, 2007.
[2] Walter Rudin. *Real Analysis: Modern Techniques and Their Applications*. John Wiley and Sons, New York, 2nd edition, 1999.