LESSON 11 - STUDY GUIDE

Abstract. Continuing the previous lesson on convolution, we will focus now on what, arguably, is its main application: approximate identities.

1. Approximate Identities.

Study material: We will continue on section **8.2 Convolution** from chapter **8 Elements of Fourier Analysis** in [1], specifically focusing on pgs. 242–246, as well as Section **1.2 Convolution and Approximate Identities**, in particular the subsection **1.2.4 Approximate Identities** in [2]. Unlike the previous lesson, for the current one I will pick the material more evenly from both sources which, I believe, complement each other on this topic.

In the previous lesson, we saw how $L^1(\mathbb{R}^n)$ is a commutative algebra with the product defined by the convolution

$$f * g(x) = \int_{\mathbb{R}^n} f(x-y)g(y)dy. \tag{1.1}$$

However, this algebra does not have an identity, for there is no function $I \in L^1(\mathbb{R}^n)$ such that

$$f * I = f.$$

In fact, such an I would be the Dirac δ, which is a distribution (actually, quite a “mild” distribution... it is simply a measure). Nevertheless, even outside the framework of the theory of distributions, there is still a very useful alternative, consisting of families or sequences of L^1 functions that, intuitively, converge to the Dirac δ and therefore approximate the identity.

Definition 1.1. *An approximate identity, or approximation of the identity, is a family of functions $k_\varepsilon \in L^1(\mathbb{R}^n)$, $\varepsilon > 0$, such that*

1. *The family is bounded in the $L^1(\mathbb{R}^n)$ norm, i.e. there is $C > 0$ such that $\|k_\varepsilon\|_{L^1(\mathbb{R}^n)} \leq C$, for all $\varepsilon > 0$.*
2. $\int_{\mathbb{R}^n} k_\varepsilon(x)dx = 1$*, for all $\varepsilon > 0$.*
3. *For any $\delta > 0$, $\lim_{\varepsilon \to 0} \int_{|x| \geq \delta} |k_\varepsilon(x)|dx = 0$.*

As mentioned above, it is also very common to consider sequences of approximate identities, in which case one takes $\{k_n\}_{n\in\mathbb{N}}$ as $n \to \infty$.

The simplest and most frequent form of obtaining approximate identities consists in rescaling a single function $\phi \in L^1(\mathbb{R}^n)$, with $\int \phi = 1$. In fact, if from one such function one defines the rescaled family

$$\phi_\varepsilon(x) = \frac{1}{\varepsilon^n}\phi\left(\frac{x}{\varepsilon}\right), \quad \text{for} \quad \varepsilon > 0, \tag{1.2}$$

then

$$\int_{\mathbb{R}^n} \phi_\varepsilon(x)dx = \int_{\mathbb{R}^n} \frac{1}{\varepsilon^n}\phi\left(\frac{x}{\varepsilon}\right) = \int_{\mathbb{R}^n} \phi(y)dy = 1,$$

Date: March 30, 2020.

by the change of variable $y = x/\varepsilon$. And, analogously, $\|\phi_\varepsilon\|_{L^1(\mathbb{R}^n)} = \|\phi\|_{L^1(\mathbb{R}^n)}$, for all $\varepsilon > 0$ (although, of course, this norm might not be 1 if ϕ is not positive). So, conditions (1) and (2) in Definition 1.1 are easily satisfied. As for condition (3), one should keep in mind that the rescaling of ϕ given by (1.2) shrinks the support of ϕ around the origin, while spiking its values, keeping the integral constant and equal to 1 (intuitively approximating a Dirac δ) as $\varepsilon \to 0$. To put it rigorously, one starts by observing that cutting off $|\phi|$ at radius r, then the corresponding integral $\int |\phi|\, \chi_{\{|x|<r\}}$ converges obviously to $\int |\phi|$ as $r \to \infty$ by the monotone convergence theorem. Thus

$$\int |\phi| - \int |\phi|\, \chi_{\{|x|<r\}} = \int |\phi| \chi_{\{|x|\geq r\}} \to 0 \quad \text{as} \quad r \to 0.$$

But then, given any fixed $\delta > 0$, we have, by a change of variables,

$$\int_{|x|\geq\delta} |k_\varepsilon(x)|dx = \int_{|x|\geq\delta} \frac{1}{\varepsilon^n} \left|\phi\left(\frac{x}{\varepsilon}\right)\right| dx = \int_{|y|\geq\frac{\delta}{\varepsilon}} |\phi(y)|dy \to 0,$$

by the preceding observation, because $r = \frac{\delta}{\varepsilon} \to \infty$ as $\varepsilon \to 0$. So condition (3) in Definition 1.1 is also satisfied and this establishes that ϕ_ε, defined by the rescaling of ϕ, is indeed an approximate identity.

The importance of approximate identities stems from the following fundamental theorem.

Theorem 1.2. *Suppose that k_ε is an approximate identity or, more generally, a family of functions that satisfies Definition 1.1, (1) and (3), with condition (2) broadened to the case*

$$(2') \qquad \int_{\mathbb{R}^n} k_\varepsilon(x)dx = a \in \mathbb{C}, \quad \textit{for all} \quad \varepsilon > 0.$$

Then we have

1. *If $f \in L^p(\mathbb{R}^n)$, with $1 \leq p < \infty$, then $f * k_\varepsilon \in L^p(\mathbb{R}^n)$ and $f * k_\varepsilon \to af$ in the $L^p(\mathbb{R}^n)$ norm, as $\varepsilon \to 0$.*
2. *If f is uniformly continuous and bounded on $\mathbb{R}^n$, then $f * k_\varepsilon$ is also uniformly continuous and bounded on $\mathbb{R}^n$ and $f * k_\varepsilon \to af$ uniformly, i.e. in the $L^\infty(\mathbb{R}^n)$ norm, as $\varepsilon \to 0$.*
3. *If $f \in L^\infty(\mathbb{R}^n)$ is continuous on an open set $\Omega \subset \mathbb{R}^n$, then $f * k_\varepsilon \in L^\infty(\mathbb{R}^n)$, is uniformly continuous on $\mathbb{R}^n$ and $f * k_\varepsilon \to af$ uniformly on compact subsets $K \subset \Omega$, as $\varepsilon \to 0$.*

Remark 1.3. *Of course the case $a = 1$, corresponding to approximate identities, is the most frequent application of this theorem. But the case $a = 0$, for example, is also frequently useful.*

Proof.

1. As $f \in L^p(\mathbb{R}^n)$, with $1 \leq p < \infty$, and $k_\varepsilon \in L^1(\mathbb{R}^n)$ then $f * k_\varepsilon \in L^p(\mathbb{R}^n)$ from Young's inequality for convolutions (Proposition 1.3 from the previous lesson).

   Now,

   $$f * k_\varepsilon(x) - af(x) = \int_{\mathbb{R}^n} f(x-y)k_\varepsilon(y)dy - \int_{\mathbb{R}^n} k_\varepsilon(y)dy f(x) = \int_{\mathbb{R}^n} \big(f(x-y) - f(x)\big)k_\varepsilon(y)dy,$$

   and we use Minkowski's integral inequality to obtain

   $$\|f * k_\varepsilon - af\|_{L^p(\mathbb{R}^n)} \leq \int_{\mathbb{R}^n} \|f(\cdot - y) - f(\cdot)\|_{L^p(\mathbb{R}^n)} |k_\varepsilon(y)|dy.$$

   To show that this quantity is arbitrarily small as $\varepsilon \to 0$, we split this integral into two pieces: in one of them, for small $|y|$, the smallness is achieved by the continuity of the translation operator in the L^p norm, $1 \leq p < \infty$, seen in the previous lesson (Theorem 1.7), while the L^1 norm of k_ε remains bounded from property (1) in Definition 1.1; for large $|y|$ these roles are switched, and while the L^p norm of the translation remains bounded, it is property (3) in Definition 1.1 that yields the smallness in this case.

So, for arbitrarily small $\zeta > 0$, pick $\delta > 0$ such that $\|f(\cdot - y) - f(\cdot)\|_{L^p(\mathbb{R}^n)} = \|\tau_y f - f\|_{L^p(\mathbb{R}^n)} < \zeta$ for all $|y| < \delta$, by the continuity of the translation map at $y = 0$ in $L^p(\mathbb{R}^n)$, $1 \le p < \infty$. Then

$$\|f * k_\varepsilon - af\|_{L^p(\mathbb{R}^n)} \le$$
$$\le \int_{|y|<\delta} \|f(\cdot - y) - f(\cdot)\|_{L^p(\mathbb{R}^n)} |k_\varepsilon(y)| dy + \int_{|y|\ge\delta} \|f(\cdot - y) - f(\cdot)\|_{L^p(\mathbb{R}^n)} |k_\varepsilon(y)| dy \le$$
$$\le \zeta \|k_\varepsilon\|_{L^1(\mathbb{R}^n)} + 2\|f\|_{L^p(\mathbb{R}^n)} \int_{|y|\ge\delta} |k_\varepsilon(y)| dy$$

and as explained above, the second term in this sum can be made arbitrarily small by property (3) in Definition 1.1 while the $\|k_\varepsilon\|_{L^1(\mathbb{R}^n)} \le C$ by property (1). This finishes the proof of (1).

(2) If f is bounded and uniformly continuous, then $f \in C_b(\mathbb{R}^n) \subset L^\infty(\mathbb{R}^n)$, where we denote by $C_b(\mathbb{R}^n)$ the space of bounded continuous functions with the supremum norm (that coincides with the L^∞ norm of the corresponding class of equivalence of functions in L^∞). Then, from Hölder's inequality for convolutions, Theorem 1.8 in the previous lesson, it follows that $f * k_\varepsilon$ is a bounded and uniformly continuous function. Incidentally, the uniform continuity of f is the same as the (uniform) continuity of the translation map in the L^∞ norm, as it corresponds to guaranteeing that
$$\sup_{x\in\mathbb{R}^n} |f(x-y) - f(x)| = \|\tau_y f - f\|_{L^\infty(\mathbb{R}^n)} \to 0, \quad \text{when} \quad y \to 0.$$
(we observed this same fact in the last lesson, when proving the uniform continuity of continuous functions of compact support, in Proposition 1.6). But then we can just repeat the proof above, in part (1), replacing the L^p norm there by the L^∞ norm. Of course, part (1) does not hold generally for $p = \infty$ because the translation map is not continuous for any function in L^∞, but the condition that f is uniformly continuous ensures that it is in this particular case.

(3) That $f * k_\varepsilon$ is bounded (L^∞) and uniformly continuous on $\mathbb{R}^n$ is again just a consequence of Hölder's inequality for convolutions, Theorem 1.8 in the previous lesson. If f is continuous on an open set Ω then, from the Heine-Cantor theorem in advanced calculus we know that f is uniformly continuous on compact subsets $K \subset \Omega$, i.e. for any $\zeta > 0$ there is $\delta > 0$ such that $\sup_{x\in K} |f(x-y) - f(x)| = \|\tau_y f - f\|_{L^\infty(K)} < \zeta$ for $|y| < \delta$ (notice that x is in K but $x - y$ need not be). The result is proved again by an argument entirely similar to the proofs of parts (1) combined with (2), but just restricting to $x \in K$ and using the continuity of the translation map with respect to the L^∞ norm on $K \subset \Omega$.

□

Recall that convergence in the L^p norm does not imply pointwise convergence, not even almost everywhere. As we have already seen, convergence in the L^p norm only ensures the existence of subsequences that converge a.e. So the previous theorem, except for the L^∞ cases (2) and (3), where some continuity of f is also demanded in the hypotheses, is still quite far from a result for pointwise convergence of approximate identities. Later in this course we will study an extremely powerful tool of harmonic analysis that is usually the predominant technique for proving pointwise convergence for sequences of functions in L^p: the Hardy-Littlewood maximal operator.

Combining approximate identities with the smoothness of convolutions (Proposition 1.4 from the previous lesson) we can improve the result on the density of C_c in L^p to show that, in $\mathbb{R}^n$, one actually has density of C_c^∞ in L^p for any open set.

Theorem 1.4. *Let $\Omega \subset \mathbb{R}^n$ be an open set. Then, $C_c^\infty(\Omega)$ is dense in $L^p(\Omega)$ for all $1 \le p < \infty$.*

Proof. Let $K_1 \subset K_2 \subset K_3 \ldots \subset \Omega$ be an increasing sequence of compact subsets of Ω such that $\cup_{j=1}^{\infty} K_j = \Omega$. Then, for any $f \in L^p(\Omega)$, $1 \leq p < \infty$, the sequence of cut-offs $f\chi_{K_j}$ converges to f in the $L^p(\Omega)$ norm, where χ_{K_j} denote the characteristic functions of the compacts K_j. In fact, we have

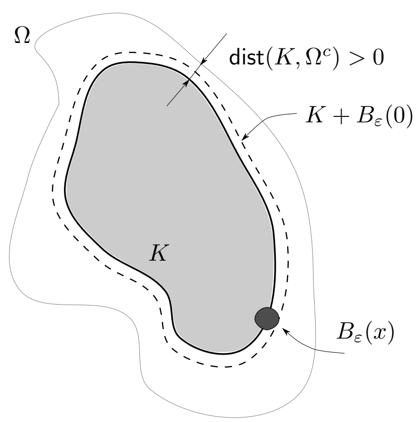
$$\int_\Omega |f(x) - f(x)\chi_{K_j}(x)|^p dx = \int_\Omega |f(x)|^p (1 - \chi_{K_j}(x)) dx \to 0,$$

by the dominated convergence theorem. So, for any $f \in L^p(\Omega)$ and arbitrarily small $\delta > 0$ we can pick a compact $K \subset \Omega$ for which $\|f - f\chi_K\|_{L^p(\Omega)} < \delta/2$.

Now that we have extracted an $L^p(\Omega)$ function $f\chi_K$, of compact support in Ω, $\delta/2$-close to the original f (but still as rough as the original f) we will use the regularity of convolutions together with approximate identities to build a $C_c^\infty(\Omega)$ function $\delta/2$-close to $f\chi_K$, therefore δ-close to f, thus establishing the density result. So, we start by choosing $\phi \in C_c^\infty(\mathbb{R}^n)$, such that $\operatorname{supp}\phi$ is contained in the unit ball around the origin $B_1(0)$ and $\int \phi = 1$ (this requires a bit of familiarity with C_c^∞ functions which I am assuming, so please review these concepts in Section 8.1 of [1] in case you lack that familiarity). It is a simple exercise of topology of subsets of $\mathbb{R}^n$ to show that the distance from a compact set K to a closed disjoint set is positive. So $\operatorname{dist}(K, \Omega^c) = \inf\{|x - y| : x \in K, y \in \Omega^c\} > 0$ and if we consider ε much smaller than this positive distance, we surely have, from property (5) of Theorem 1.1 in the last lecture,

$$\operatorname{supp}(f\chi_K * \phi_\varepsilon) \subset \overline{K + B_\varepsilon(0)} \subset \Omega,$$

where ϕ_ε is the rescaled version of ϕ as in (1.2).



In other words, the support of $f\chi_K * \phi_\varepsilon$ is contained in the closure of the ε-neighborhood of K. On the other hand, because K is compact, thus of finite measure, we have $f\chi_K \in L^1(\Omega)$ from the inclusion of L^p spaces on finite measure domains. And from the regularity result for convolutions, Proposition 1.4 in the last lecture, we can finally conclude that $f\chi_K * \phi_\varepsilon \in C_c^\infty(\Omega)$.

To conclude the argument, we know from the approximation of the identity property of ϕ_ε, that $f\chi_K * \phi_\varepsilon \to f\chi_K$ in the $L^p(\Omega)$ norm, as $\varepsilon \to 0$. So we can pick a particular $f\chi_K * \phi_\varepsilon \in C_c^\infty(\Omega)$ such that $\|f\chi_K - f\chi_K * \phi_\varepsilon\|_{L^p(\Omega)} < \delta/2$ which then satisfies

$$\|f - f\chi_K * \phi_\varepsilon\|_{L^p(\Omega)} \leq \|f - f\chi_K\|_{L^p(\Omega)} + \|f\chi_K - f\chi_K * \phi_\varepsilon\|_{L^p(\Omega)} < \delta,$$

as desired. □

We will finish this section with another very useful and related result: the existence of smooth cut-off functions for compact sets. The existence of such cut-off, or bump, functions is an essential ingredient, for example, in the construction of smooth partitions of unity (their existence can also be proved without convolutions, but here it becomes a simple application of the results developed so far).

Theorem 1.5. **(C^∞ Urysohn Lemma - Existence of smooth cut-off functions)** *Let $K \subset \Omega$ be a compact subset of the open set Ω. Then, there exists $f \in C_c^\infty$, with $0 \leq f \leq 1$ such that $f(x) = 1$ for all $x \in K$ and $\operatorname{supp} f \subset \Omega$.*

Proof. As before, $\operatorname{dist}(K, \Omega^c) = \inf\{|x-y| : x \in K,\ y \in \Omega^c\} > 0$. Let us denote this distance by D. Considering $\delta = D/3$, say, and the δ-neighborhood of the compact set K

$$K_\delta = \{x \in \Omega : |x-y| \leq \delta \quad \text{for some} \quad y \in K\},$$

we take the characteristic function χ_{K_δ}.

Choose now a non-negative $\phi \in C_c^\infty$ with $\int \phi = 1$ and $\operatorname{supp} \phi \subset B_1(0)$ as in the previous proof. Again consider the rescaled ϕ_ε and make $\varepsilon = D/10$, for example. Then, from the property about the supports of convolutions, we have that

$$\operatorname{supp} \chi_{K_\delta} * \phi_\varepsilon \subset \overline{K_\delta + B_\varepsilon(0)} \subset K_{\delta+\varepsilon} \subset \Omega,$$

and it easy to check that, for $x \in K$

$$\chi_{K_\delta} * \phi_\varepsilon(x) = \int_{K_\delta} \phi_\varepsilon(x-y)dy = 1.$$

So $f = \chi_{K_\delta} * \phi_\varepsilon$ satisfies the required properties and this finishes the proof. □

References

[1] Gerald B. Folland, *Real Analysis, Modern Techniques and Applications,* 2nd Edition, John Wiley & Sons, 1999.
[2] Loukas Grafakos, *Classical Fourier Analysis,* 3rd Edition, Sp ringer, Graduate Texts in Mathematics 249, 2014.