LESSON 14 - STUDY GUIDE

Abstract. In this lesson we will start looking at some decay properties of the Fourier coefficients, and how they relate to the regularity of the functions. We will also begin to address the general issue of convergence of Fourier series, by considering its symmetric partial sums and introducing the Dirichlet kernel to which they are intimately related.

1. Fourier series: definitions and basic properties of Fourier coefficients.

<u>**Study material:**</u> We will be expanding some of the ideas in section **2 - Summability in Norm and Homogeneous Banach Spaces** from chapter **I - Fourier Series on** $\mathbb{T}$, corresponding to pgs. 8–16 in the second edition [1] and pgs. 9–17 in the third edition [2] of Katznelson's book. But we will only complete this section in the following lesson.

In the previous lesson we presented the first definitions and basic properties of Fourier series and coefficients. In particular we saw that the sequence of Fourier coefficients of a function in $L^1(\mathbb{T})$ is always bounded. Just this fact by itself is far from characterizing all sequences of complex numbers that correspond to Fourier coefficients, as the set $\mathcal{F}(L^1(\mathbb{T}))$ is, in reality, a very small subset of $l^\infty(\mathbb{Z})$. As a matter of fact, a prevailing theme in Fourier analysis is the attempt to characterize Fourier transforms[1] and to try to deduce, as accurately as possible, the properties of the original functions from them. Naturally, the first problem that most obviously comes to mind, in that regard, is the issue of uniqueness, i.e. whether or not each function gives rise to its own unique sequence of Fourier coefficients, not shared by any other function. A sort of unique fingerprint of a function in frequency space. In other words, this is the problem of injectivity of the Fourier transform operator. The other question, that immediately ensues, is whether the original function can then be reconstructed from its Fourier transform. This, now, is the problem of finding the inverse Fourier transform operator. Of course the inversion problem is a stronger one, because injectivity follows from it as no two different functions could then be reconstructed from the same Fourier transform. If both these questions are answered positively then, necessarily, all the properties and information about the original function should somehow be encoded within its Fourier transform. And the issue then becomes how to read off these properties directly from it. In this lesson we will start looking at some of these problems.

We begin by recalling that, in $\mathbb{T}$, the $L^p(\mathbb{T})$ spaces are nested, as p decreases, and thus $L^1(\mathbb{T})$ contains them all. At the opposite end, we can identify the continuous functions on $\mathbb{T}$, which are necessarily bounded because $\mathbb{T}$ is compact, with their classes of equivalence in $L^\infty(\mathbb{T})$ (because no two different continuous functions can be equal almost everywhere with respect to the Lebesgue measure). And, contained in the space of continuous functions, we can then consider also the higher regularity spaces of k-differentiable functions. We should therefore think of a hierarchy of spaces of decreasing regularity

$$C^\infty(\mathbb{T}) \subset C^k(\mathbb{T}) \subset C(\mathbb{T}) \subset L^\infty(\mathbb{T}) \subset L^p(\mathbb{T}) \subset L^1(\mathbb{T}).$$

Date: May 4, 2020.

[1] It would probably be more accurate and rigorous to denote the operator $\mathcal{F}$ by *Fourier transformation*, while calling its output for a particular function the *Fourier transform* of that function. But following the common practice - which, for that matter, is also used for other functional operators, like the derivative for example - we somewhat abusively call both, the operator as well as its image, *Fourier transform*.

Differentiable functions on $\mathbb{T}$ should, of course, be regarded as differentiable 2π-periodic functions on $\mathbb{R}$. This is one advantage of doing Fourier series on $\mathbb{T}$, when compared to the Fourier transform on $\mathbb{R}^n$: the underlying group being compact, hence with finite total measure, implies that $L^1(\mathbb{T})$, on which the Fourier transform is naturally defined, is the largest space and contains all the others under consideration. So we are automatically equipped with an all-inclusive arena just by considering the space on which it makes sense to define the integrals. On $\mathbb{R}^n$, however, although the Fourier transform is also naturally defined on $L^1(\mathbb{R}^n)$, the $L^p(\mathbb{R}^n)$ spaces are not contained in each other and, therefore, one needs to find a larger space of objects that includes them all, on which to work in a unified manner. And that is why we are led to consider the space of distributions in that case.

Nevertheless, it should also be pointed out that "rougher" spaces of objects could be equally considered on $\mathbb{T}$, most notably measures and distributions too. As a quick useful example, it is easy to see how the definitions that we have seen before for functions can be obviously extended to the space of Borel measures on $\mathbb{T}$, that we will denote by $\mathcal{M}(\mathbb{T})$. In particular, the functions $f \in L^1(\mathbb{T})$ themselves can be identified with the measures $f(t)dt/2\pi$ and, from the Radon-Nikodym theorem, we actually know that all absolutely continuous Borel measures with respect to the Lesbesgue measure correspond to such $L^1(\mathbb{T})$ functions in this manner. This way we can think of the inclusion $L^1(\mathbb{T}) \subset \mathcal{M}(\mathbb{T})$ as identifying $L^1(\mathbb{T})$ with the subspace of absolutely continuous measures within the general Borel measures, that now becomes an even larger space of less regular objects on which to think about Fourier series. As $\mathbb{T}$ is compact and the measures $\mu \in \mathcal{M}(\mathbb{T})$ are Borel, they have finite total variation $|\mu|(\mathbb{T}) < \infty$ and this is the natural norm on $\mathcal{M}(\mathbb{T})$. In particular, for $f \in L^1(\mathbb{T})$, we have $\|f(t)dt/2\pi\|_{\mathcal{M}(\mathbb{T})} = |f(t)dt/2\pi|(\mathbb{T}) = \frac{1}{2}\int_{\mathbb{T}} |f(t)|dt = \|f\|_{L^1(\mathbb{T})}$ and therefore the identification $f \mapsto f(t)dt/2\pi$ between functions in $L^1(\mathbb{T})$ and the subspace of absolutely continuous measures in $\mathcal{M}(\mathbb{T})$ is actually an isometric embedding. Finally, as $e^{-int} \in L^1(\mathbb{T}, \mu)$, the Fourier coefficients of a measure $\mu \in \mathcal{M}(\mathbb{T})$ can simply be defined as

$$\hat{\mu}(n) = \int_{\mathbb{T}} e^{-int} d\mu(t),$$

with the bound

$$|\hat{\mu}(n)| \leq |\mu|(\mathbb{T}) = \|\mu\|_{\mathcal{M}(\mathbb{T})}, \quad \text{for all} \quad n \in \mathbb{Z}.$$

This obviously coincides with the previous definition for $f \in L^1(\mathbb{T})$ when the measure is $f(t)dt/2\pi$.

One important example to keep in mind is the Dirac-δ measure at the origin, $\delta \in \mathcal{M}(\mathbb{T})$, which is defined by $\delta(E) = 1$, if $0 \in E$, for $E \subset \mathbb{T}$, and $\delta(E) = 0$, if $0 \notin E$. Then

$$\hat{\delta}(n) = \int_{\mathbb{T}} e^{-int} d\delta(t) = 1,$$

for all $n \in \mathbb{Z}$. The Dirac-δ measure is singular, with respect to the Lebesgue measure on $\mathbb{T}$ and cannot be given by any function $f \in L^1(\mathbb{T})$.

So, although we also have $\mathcal{F} : \mathcal{M}(\mathbb{T}) \to l^\infty(\mathbb{Z})$ we will see now that a stronger property is true for $f \in L^1(\mathbb{T})$: the sequence of Fourier coefficients actually converges to zero as $|n| \to \infty$.

Theorem 1.1. (Riemann-Lebesgue lemma) *Let $f \in L^1(\mathbb{T})$. Then, its Fourier coefficients satisfy*

$$\lim_{|n|\to\infty} \hat{f}(n) = 0.$$

Proof. We saw in Lesson 11, Theorem 1.4, that for $\Omega \in \mathbb{R}^n$ open, the set $C_c^\infty(\Omega)$ is dense in $L^1(\Omega)$. We can easily adapt this result to show that $C^\infty(\mathbb{T})$ (whose functions always have compact support now because $\mathbb{T}$ is compact) is dense in $L^1(\mathbb{T})$. In fact, one can start by considering the open interval $]0, 2\pi[$ to conclude that the set of functions in $C_c^\infty(]0, 2\pi[)$ is dense in $L^1(]0, 2\pi[) = L^1([0, 2\pi[)$ and identify the latter with $L^1(\mathbb{T})$.

So, consider first $g \in C^1(\mathbb{T})$. Then, integrating by parts, we have for $n \neq 0$,

$$\hat{g}(n) = \frac{1}{2\pi}\int_0^{2\pi} g(t)e^{-int}dt = \frac{1}{2\pi}\int_0^{2\pi} \frac{g'(t)}{in}e^{-int}dt = \frac{\widehat{(g')}(n)}{in},$$

where we used the fact that $g \in C^1(\mathbb{T})$ is 2π-periodic, as a function on $\mathbb{R}$, and thus $g(0) = g(2\pi)$. We also have that g' is continuous on $\mathbb{T}$ which is, therefore, an $L^1(\mathbb{T})$ function. So, from the properties of the Fourier coefficients seen in the last lesson, we have $|\widehat{(g')}(n)| \leq \|g'\|_{L^1(\mathbb{T})}$ for all $n \in \mathbb{Z}$. And this then shows that

$$|\hat{g}(n)| \leq \frac{\|g'\|_{L^1(\mathbb{T})}}{n} \to 0, \quad \text{as} \quad n \to \infty. \tag{1.1}$$

To conclude the proof, we use the fact that $C^\infty(\mathbb{T})$, and thus also $C^1(\mathbb{T})$, is dense in $L^1(\mathbb{T})$, to pick a sequence $g_j \in C^1(\mathbb{T})$ such that $g_j \to f$ in the $L^1(\mathbb{T})$ norm. Then, from Corollary 1.5 in the last lesson, we know that $\widehat{g_j}(n) \to \hat{f}(n)$ as $j \to \infty$ uniformly in $n \in \mathbb{Z}$. So, from (1.1) for all of the $\widehat{g_j}$, this implies also that $\hat{f}(n) \to 0$ when $n \to \infty$. □

Observe that the proof of the Riemann-Lebesgue Lemma that we have just presented hinges essentially on just two ingredients: on the one hand, the fact that smooth functions are dense in $L^1(\mathbb{T})$; on the other hand, that, for smooth functions, the convergence to zero of the Fourier coefficients is a simple consequence of integration by parts.

The density property does not hold for measures, and as we saw with the Dirac-δ example, its Fourier coefficients indeed do not decay to zero at infinity. As for the convergence to zero of the Fourier coefficients of differentiable functions, it is absolutely crucial to take advantage of the oscillations of the exponentials, through the process of integration by parts, to conclude it. Had we applied absolute values inside the integral right at the start, we would have killed the oscillations, and only the constant bound $\|g\|_{L^1(\mathbb{T})}$ would be achieved.

Integrals of the form

$$I(\lambda) = \int e^{i\lambda\phi(x)} f(x)dx, \tag{1.2}$$

are called *oscillatory integrals*, and they permeate the whole of harmonic analysis, in different forms and circumstances, of which, of course, the integrals defining the Fourier transform are the prime example. In the general form (1.2), the function ϕ is usually called the *phase* of the oscillatory exponential, while $f(x)$ is called the amplitude. By using an integration by parts argument, generalizing the one that we used in the proof of the Riemann-Lebesgue Lemma, it is easy to show that, if $\phi' \neq 0$, then $I(\lambda) \to 0$ when $\lambda \to \infty$. So, any significantly large contribution to the asymptotic values of $I(\lambda)$, as $\lambda \to \infty$, can only arise from points where $\phi' = 0$ (we can use cut-off functions to isolate these points of the domain from the remaining ones, and correspondingly split the integrals into regions where either ϕ' is zero or not). This is called the *principle of stationary phase*: the relevant contributions to the asymptotic values of an oscillatory integral result from the points where the phase is stationary which, if nonexistent, implies that the integral vanishes as $\lambda \to \infty$. One can intuitively understand this principle by imagining that, if $\phi' \neq 0$, then, as $\lambda \to \infty$, the function $e^{i\lambda\phi(x)}$ oscillates very rapidly around x so that, for a reasonably regular amplitude f around that point, the "negative and positive" contributions to the integral resulting from the oscillations over a slowly varying f around x, cancel each other and make the integral vanish.

I insist again on the observation that an elementary absolute value estimate inside the integral erases the imaginary exponential, and this whole subtle picture cannot then be seen: the key to these estimates is really to carefully exploit the oscillations as best as possible. Many of the topics in more advanced harmonic analysis revolve around the subject of estimating oscillatory integrals, or their applications:

Van-der-Corput Lemma, Fourier transform of measures supported on surfaces, restriction conjecture, Strichartz estimates, etc. See Stein's "Harmonic Analysis" book [3] if you are curious and would like to look further into these topics.

Returning to the proof of the Riemann-Lebesgue Lemma, we can easily see that, had we used $C^k(\mathbb{T})$ or even $C^\infty(\mathbb{T})$ functions, instead of $C^1(\mathbb{T})$, we could successively repeat the iteration by parts computation, to obtain even faster decays of the Fourier coefficients at infinity. Actually, one of the central themes in harmonic analysis is the relation between derivatives, or, more generally, regularity of functions, and the Fourier transform. That is the the reason why Fourier analysis tools are often essential in the study of fine properties of functions (like in microlocal analysis or in the theory of spaces of functions definied in terms of different levels of regularity, like Sobolev and Besov spaces). A first, simple, but fundamental result in that direction is the following.

Proposition 1.2. *Let $f \in C^k(\mathbb{T})$. Then*

$$\widehat{f^{(k)}}(n) = (in)^k \hat{f}(n), \tag{1.3}$$

and, for $n \neq 0$,

$$|\hat{f}(n)| \leq \frac{\|f^{(k)}\|_{L^1(\mathbb{T})}}{n^k}. \tag{1.4}$$

Proof. We could start from the Fourier transform of f and integrate by parts k-times to obtain the Fourier transform of $f^{(k)}$, in the same manner as was done in the proof of the Riemann-Lebesgue lemma for the first derivative. Or, in the opposite direction, start from the Fourier transform of $f^{(k)}$ and integrate by parts k-times to reach the Fourier transform of f, with the identity (1.3). Let us do the latter this time

$$\widehat{f^{(k)}}(n) = \frac{1}{2\pi}\int_{\mathbb{T}} f^{(k)}(t)e^{-int}dt = \frac{in}{2\pi}\int_{\mathbb{T}} f^{(k-1)}(t)e^{-int}dt = \ldots = \frac{(in)^k}{2\pi}\int_{\mathbb{T}} f(t)e^{-int}dt = (in)^k\hat{f}(n).$$

Obviously, (1.4) follows from (1.3) by using the estimate $|\widehat{f^{(k)}}(n)| \leq \|f^{(k)}\|_{L^1(\mathbb{T})}$. □

Identity (1.3) (as well as its siblings for the Fourier transform in $\mathbb{R}^n$) is one of the most basic and crucial properties in Fourier analysis, as it shows that the Fourier transform maps derivatives of the original function to multiplication of the Fourier coefficients by powers of the frequency. More generally, a differential operator on $\mathbb{T}$ corresponds then to multiplication by a polynomial on the frequency side. This is precisely why harmonic analysis is such an important tool in the study of differential equations, for it is typically much easier to solve a polynomial equation on the frequency side, after applying the Fourier transform, than it is to solve the original purely differential equation.

On the other hand (1.4) shows that, the smoother the function is, the faster is the decay of the Fourier coefficients as $n \to \infty$. At one extreme we have rough measures, like the Dirac-δ, whose Fourier transform is constant and does not even decay, while at the other extreme, with smooth $C^\infty(\mathbb{T})$ functions, the Fourier transform decays faster than any power of the frequency. This again is another fundamental and pervasive property in harmonic analysis: regularity of the functions translates into decay of the Fourier transform.

Notice, however, that (1.4), although necessary, is not a sufficient property for smoothness of the function. For example, the characteristic function of the interval $[-1, 1]$, regarded as a function on $\mathbb{T}$, has Fourier coefficients

$$\widehat{\chi_{[-1,1]}}(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \chi_{[-1,1]}(t)e^{-int}dt = \frac{1}{2\pi}\int_{-1}^{1} e^{-int}dt = -\frac{1}{2i\pi n}e^{-int}\Big|_{-1}^{1} = \frac{\sin(nt)}{\pi n},$$

which decay as $\sim \dfrac{1}{n}$, although $\chi_{[-1,1]}$ is not C^1, not even continuous, on $\mathbb{T}$.

If we finally start looking at the convergence of Fourier series, for general $f \in L^1(\mathbb{T})$

$$S[f] \sim \sum_{n=-\infty}^{\infty} \hat{f}(n) e^{int},$$

then, the previous discussion about the decay as $n \to \infty$ of the Fourier coefficients makes it look as if we should generally not expect this series to converge. In particular, if we demand absolute convergence, i.e.

$$\sum_{n=-\infty}^{\infty} |\hat{f}(n)| < \infty,$$

then it seems as if we can only guarantee it for functions at least twice continuously differentiable, for then $|\hat{f}(n)| \leq C/n^2$, by (1.4). On the other hand, absolute convergence immediately implies uniform convergence (by the Weierstrass M-test, from advanced calculus) and therefore the resulting function, arising from the sum of the series, would be necessarily continuous. So, absolute convergence looks like a hopelessly high bar for any $f \in L^1(\mathbb{T})$ which is not, at least, continuous on the whole circle $\mathbb{T}$. And we will see a few lessons ahead that this is indeed the case.

We could, of course, consider only pointwise conditional convergence (absolute convergence at a single point $t \in \mathbb{T}$ obviously implies absolute and uniform convergence on the whole $\mathbb{T}$). But, as is also very well known from advanced calculus, series which converge conditionally have sums which are highly unstable to changes in the order of the terms. So this is not a robust form of summation to be considered. But pointwise convergence is indeed probably the most intuitive form of convergence and the one that spontaneously comes to mind when thinking about recovering a function from its Fourier series. And, although it has been a constant central object of study, from the first attempts by Fourier and Dirichlet, to the impressive theorem of Carleson, in 1966, the reality is that any other form of convergence, or more generally, any method that allows for the reconstruction of a function from its Fourier coefficients is equally significant. With the development of the Lebesgue theory of integration, at the beginning of the 20th century, new forms of looking at the convergence of Fourier series came to the fore: convergence in the L^p norm and summability methods.

From here on, when we refer to the *convergence of Fourier series* we will always consider it as meaning the convergence of symmetric partial sums, in some norm or at fixed points $t \in \mathbb{T}$

$$S_N[f](t) = \sum_{n=-N}^{N} \hat{f}(n) e^{int}, \tag{1.5}$$

because they are equivalent to the partial sums for the Fourier series in real form, as originally written by Fourier,

$$\frac{a_0}{2} + \sum_{n=1}^{N} a_n \cos(nt) + b_n \sin(nt).$$

The partial sum $S_N[f]$ is evidently a trigonometric polynomial of degree $\leq N$ and with coefficients given by the truncated Fourier coefficients of f, $\hat{f}(n)$ for $-N \leq n \leq N$. One can think of it as the product of the full sequence of Fourier coefficients of f, $\{\hat{f}(n)\}_{n\in\mathbb{Z}}$, with the characteristic sequence over the integers $-N \leq n \leq N$, i.e. the Fourier coefficients of the trigonometric polynomial $\sum_{-N\leq n\leq N} e^{int}$. But, from Proposition 1.7 in the previous lesson, we know that this is exactly the result of the convolution of f and $\sum_{-N\leq n\leq N} e^{int}$. So we can conclude that:

$$S_N[f](t) = \sum_{n=-N}^{N} \hat{f}(n) e^{int} = f * \sum_{n=-N}^{N} e^{int}.$$

This can also be shown by a simple direct computation, as

$$= \sum_{n=-N}^{N} \hat{f}(n) e^{int} = \sum_{n=-N}^{N} \left(\frac{1}{2\pi} \int_{\mathbb{T}} f(s) e^{-ins} ds \right) e^{int} =$$

$$= \frac{1}{2\pi} \int_{\mathbb{T}} f(s) \left(\sum_{n=-N}^{N} e^{in(t-s)} \right) ds = f * \sum_{n=-N}^{N} e^{int}.$$

In other words, the symmetric partial sum of the Fourier series of $f \in L^1(\mathbb{R})$ is simply the convolution of f and the trigonometric polynomial with unitary coefficients, truncated between frequencies $-N$ and N. Our goal is to study the behavior of this convolution as the interval of frequencies captured by the truncated polynomial widens to infinity. This polynomial is a central object, therefore, in the study of the convergence of Fourier series.

Definition 1.3. *The trigonometric polynomial of degree N, with all coefficients equal to 1, is called the* Dirichlet kernel *and denoted by D_N,*

$$D_N(t) = \sum_{n=-N}^{N} e^{int} = \frac{\sin(N + \frac{1}{2})t}{\sin \frac{t}{2}}.$$

(the last identity is a simple exercise in the summation of geometric sequences.)

Putting everything together, we therefore conclude that

$$S_N[f](t) = f * D_N(t),$$

which should immediately remind us of Lesson 11 and approximate identities. Unfortunately, however, the Dirichlet kernel *is not* an approximate identity, and this is exactly the reason why the convergence of Fourier series, either in the L^p norm, or pointwise, is so difficult and very often does not work.

In the following lesson we will see how to bypass this difficulty, and reconstruct the function f from its Fourier coefficients, by simply using, not the Dirichlet kernel, but any other sequence of polynomials that constitute an approximate identity instead. Of course, the resulting objects will not correspond to the partial sums of the Fourier series, but nevertheless will allow us to recover f uniquely from the sequence $\{\hat{f}(n)\}_{n\in\mathbb{Z}}$. Those are called summability methods.

References

[1] Yitzhak Katznelson *An Introduction to Harmonic Analysis,* 2nd Edition, Dover Publications, 1976.
[2] Yitzhak Katznelson *An Introduction to Harmonic Analysis,* 3rd Edition, Cambridge University Press, 2004.
[3] Elias Stein, *harmonic Analysis,* Princeton Univ. Press, 1993.