Lecture 1

Introduction to Fourier Analysis

Jan 7, 2005
Lecturer: Nati Linial
Notes: Atri Rudra & Ashish Sabharwal

1.1 Text

The main text for the first part of this course would be

- T. W. Körner, *Fourier Analysis*

The following textbooks are also "fun"

- H. Dym and H. P. Mckean, *Fourier Series and Integrals*.
- A. Terras, *Harmonic Analysis on Symmetric Spaces and Applications, Vols. I, II*.

The following text follows a more terse exposition

- Y. Katznelson, *An Introduction to Harmonic Analysis*.

1.2 Introduction and Motivation

Consider a vector space V (which can be of finite dimension). From linear algebra we know that at least in the finite-dimension case V has a basis. Moreover, there are more than one basis and in general different bases are the same. However, in some cases when the vector space has some additional structure, some basis might be preferable over others. To give a more concrete example consider the vector space $V = \{f : X \to \mathbb{R} \text{ or } \mathbb{C}\}$ where X is some universe. If $X = \{1, \cdots, n\}$ then one can see that V is indeed the space $\mathbb{R}^n$ or $\mathbb{C}^n$ respectively in which case we have no reason to prefer any particular basis. However, if X is an abelian group[1] there may be a reason to prefer a basis $\mathcal{B}$ over others. As an example, let us consider $X = \mathbb{Z}/n\mathbb{Z}$, $V = \{(y_0, \cdots, y_n) | y_i \in \mathbb{R}\} = \mathbb{R}^n$. We now give some scenarios (mostly inspired by the engineering applications of Fourier Transforms) where we want some properties for $\mathcal{B}$ aka our "wish list":

[1] An abelian group is given by $\langle S, +\rangle$ where S is the set of elements which is closed under the commutative operation $+$. Further there exists an identity element 0 and every element in S has an inverse.

1. Think of the elements of $\mathbb{Z}/n\mathbb{Z}$ as time units and let the vector $\boldsymbol{y} = (y_1, \cdots, y_n)$ denote some measurements taken at the different time units. Now consider another vector $\boldsymbol{z} = \{z_1, \cdots, z_n\}$ such that for all i, $z_i = y_{i+1 \mod (n)}$. Note that the vectors $\boldsymbol{y}$ and $\boldsymbol{z}$ are different though from the measurement point of view they are not much different– they correspond to the same physical situation when our clock was shifted by one unit of time. Thus, with this application in mind, we might want to look for a basis for V such that the representation of $\boldsymbol{z}$ is "closely related" to that of $\boldsymbol{y}$.

2. In a setting more general than the previous one, if for $f : X \to \mathbb{R}$, a given member of V and $a \in X$, $g : X \to \mathbb{R}$ be such that $g(x) = f(x + a)$ then we would like to have representations of f and g being "close" in $\mathcal{B}$ for any f and a. Note that in the previous example x corresponds to the index i and $a = 1$.

3. In situations where derivatives (or discrete analogues thereof are well-defined), we would like f' to have a representation similar to that of f.

4. In real life, signals are never nice and smooth but suffer from noise. One way to reduce the noise is to "average-out" the signal. As a concrete example let the signal samples be given by $f_0, \cdots, f_{n-1}$ then the smoothened out signal could be given by the samples $g_i = \frac{1}{4}f_{i-1} + \frac{1}{2}f_i + \frac{1}{4}f_{i+1}$. Define $g_{-1} = \frac{1}{4}, g_0 = \frac{1}{2}, g_1 = \frac{1}{4}$. We now look at an new operator: *convolution* which is defined as follows. Let g be f convolved where $h = f * g$ and $h(x) = \sum_y f(y)g(x - y)$. So another natural property of $\mathcal{B}$ is that the representation of $f * g$ should be related to that of f and g.

Before we go ahead, here are some frequently used instantiations of X:

- $X = \mathbb{Z}/n\mathbb{Z}$. This is the Discrete Fourier Transform (DFT).
- $X = \mathbb{T} = \{\text{Real Numbers mod } 1, \text{addition mod } 1\} \cong \{e^{i\theta}, \text{multiplication}\}$. The isomorphism exists because multiplication of elements in the second group is the same as addition mod 2π of the angles θ.

  This is the most classical case of the theory, covered by the book *Trigonometric Polynomials* by Zygmund.
- $X = (\mathbb{R}, +)$. This is the Real Fourier Transform. In this case, in order to get meaningful analysis, one has to restrict the family of functions $f : X \to \mathbb{R}$ under consideration e.g. ones with converging integrals or those with compact support. The more general framework is that of Locally compact Abelian groups.
- $X = \{0, 1\}^n$ where the operations are done mod 2. Note that $\{f : X \to \{0, 1\}\}$ are simply the boolean functions.

1.3 A good basis

As before let X be an abelian group and define the vector space $V = \{f : X \to \mathbb{R}\}$.

Definition 1.1. The *characters* of X is the set $\{\chi : X \to \mathbb{C} | \chi \text{ is a homomorphism}\}$.

By homomorphism we mean that the following relationship holds: $\chi(x+y) = \chi(x) \cdot \chi(y)$ for any $x, y \in X$. As a concrete example, $X = \mathbb{Z}/n\mathbb{Z}$ has n distinct characters and the jth character is given by $\chi_j(x) = \omega^{jx}$ for any $x \in X$ where $\omega = e^{2\pi i/n}$.

We now state a general theorem without proof (we will soon prove a special case):

Theorem 1.1. *Distinct characters (considered as functions from $X \to \mathbb{C}$) are orthogonal*[2].

We now have the following fact:

Fact 1.1. If X is a finite abelian group of n elements, then X has n distinct characters which form an orthogonal basis for $V = \{f : X \to \mathbb{R}\}$.

Consider the special case of $X = \mathbb{Z}/n\mathbb{Z}$. We will show that the characters are orthogonal. Recall that in this case, $\chi_j(x) = \omega^{jx}$. By definition, $\langle \chi_j, \chi_k \rangle = \sum_{x=0}^{n-1} \omega^{jx}\omega^{-kx} = \sum_{x=0}^{n-1} \omega^{(j-k)x}$. If $j = k$ then each term is one and the inner product evaluates to n. If $j \neq k$, then summing up the geometric series, we have $\langle \chi_j, \chi_k \rangle = \frac{(\omega^{j-k})^n - 1}{\omega^{j-k}} = 0$. The last equality follows from the fact that $\omega^n = 1$.

We will take a quick detour and mention some applications where Fourier Analysis has had some measure of success:

- Coding Theory. A code $\mathcal{C}$ is a subset of $\{0,1\}^n$ where we want each element to be as far as possible from each other (where far is measured in terms of the hamming distance). We would like $\mathcal{C}$ to be as large as possible while keeping the distance as large as possible. Note that these are two opposing goals.

- Influence of variables on boolean functions. Say you have an array of sensors and there is some function which computes an answer. If a few of the sensors fail then answers should not change: in other words we need to find functions that are not too influenced by their variables.

- Numerical Integration/ Discrepancy. Say you want to integrate over some domain Ω. Of course one cannot find the exact integral if one does not have have an analytical expression of the function. So one would sample measurements at some discrete points and try and approximate the integral. Suppose that we further know that certain subdomains of Ω are significant for the computation. The main question is how to spread n points in Ω such that every “significant” region is sampled with the “correct” number of points.

1.4 A Rush Course in Classical Fourier Analysis

Let $X = \mathbb{T} = \left(\left\{e^{i\theta} \mid 0 \leq \theta < 2\pi\right\}, \text{multiplication}\right)$. Let $f : \mathbb{T} \to \mathbb{C}$, which can alternatively be thought of as a periodic function $f : \mathbb{R} \to \mathbb{C}$. What do characters of X look like?

There are infinitely many characters and each is a periodic function from $\mathbb{R}$ to $\mathbb{C}$. In fact, every character of X is a function $\chi : \mathbb{X} \to \mathbb{T}$, i.e. $\chi : \mathbb{T} \to \mathbb{T}$. Being a homomorphism, it must also satisfy $\chi(x.y) =$

[2] There is natural notion of inner-product among functions $f, g : X \to \mathbb{R}$. $\langle f, g \rangle = \sum_{x \in X} f(x)g(x)$ in the discrete case and $\langle f, g \rangle = \int f(x)g(x)\, dx$ in the continuous case. If the functions maps into $\mathbb{C}$, then $g(x)$ is replaced by its conjugate $\overline{g}(x)$ in the expressions. Finally f and g are *orthogonal* if $\langle f, g \rangle = 0$

$\chi(x).\chi(y)$. This implies that the only continuous characters of X are $\chi_k(x) = x^k, k \in \mathbb{Z}$. Note that if $k \notin \mathbb{Z}$, then x^k can have multiple values, discontinuities, etc. It is an easy check to see that $\langle \chi_k, \chi_l \rangle = \delta_{kl}$:

$$\begin{aligned}
\langle \chi_k, \chi_l \rangle &= \frac{1}{2\pi} \int_{\mathbb{T}} \chi_k(x) \, \overline{\chi_l(x)} \, dx \\
&= \frac{1}{2\pi} \int_{\mathbb{T}} x^k x^{-l} \, dx \\
&= \frac{1}{2\pi} \int_{\mathbb{T}} x^{k-l} dx \\
&= \frac{1}{2\pi} \int_0^{2\pi} e^{i\theta(k-l)} \, d\theta \\
&= \begin{cases} 1 & \text{if } k = l \\ 0 & \text{if } k \neq l \end{cases} \\
&= \delta_{kl}
\end{aligned}$$

How do we express a given $f : \mathbb{T} \to \mathbb{C}$ in the basis of the characters of X? Recall that if V is a finite dimentional vector space over a field $\mathbb{F}$ with an inner product and $u_1, \ldots, u_n$ is an orthonormal basis for V, then every $f \in V$ can be expressed as

$$f = \sum_{j=1}^{n} a_j u_j, \quad a_j \in \mathbb{F}, a_j = \langle f, u_j \rangle \tag{1.1}$$

We would like to obtain a similar representation of f in the basis of the characters $\chi_k, k \in \mathbb{Z}$.

Definition 1.2 (Fourier Coefficients). For $r \in \mathbb{Z}$, the r^{th} Fourier coefficient of f is

$$\hat{f}(r) = \frac{1}{2\pi} \int_{\mathbb{T}} f(t) \, e^{-irt} \, dt$$

The analogue of Equation 1.1 now becomes

$$S_n(f,t) = \sum_{r=-n}^{n} \hat{f}(r) \, e^{irt}, \quad \text{does } \lim_{n \to \infty} S_n(f,t) = f(t)? \tag{1.2}$$

Here $\hat{f}(r)$ replaces a_j and $\chi_r(e^{it})$ replaces u_j in Equation 1.1. In a dream world, we would like to ask whether $\sum_{r=-\infty}^{\infty} \hat{f}(r) \, e^{irt} \stackrel{?}{=} f(t)$ holds. We are, however, being more careful and asking the question by making this sum go from $-n$ to n and considering the limit as $n \to \infty$.

1.4.1 Notions of Convergence

Before attempting to answer the question of representation of f in terms of its Fourier coefficients, we must formalize what it means for two functions defined over a domain A to be "close". Three commonly studied notions of distance between functions (and hence of convergence of functions) are as follows.

L_∞ **Distance:** $||f - g||_\infty = \sup_{x \in A} |f(x) - g(x)|$. Recall that convergence in the sense of L_∞ is called *uniform convergence*.

L_1 **Distance:** $||f-g||_1 = \int_A |f(x)-g(x)|\, dx$

L_2 **Distance:** $||f-g||_2 = \sqrt{\int_A |f(x)-g(x)|^2\, dx}$

In Fourier Analysis, all three measures of proximity are used at different times and in different contexts.

1.4.2 Fourier Expansion and Fejer's Theorem

The first correct proof (under appropriate assumptions) of the validity of Equation 1.2 was given by Dirichlet:

Theorem 1.2 (Dirichlet). *Let $f : \mathbb{T} \to \mathbb{C}$ be a continuous function whose first derivative is continuous with the possible exception of finitely many points. Then Equation 1.2 holds for every $t \in \mathbb{T}$ at which f is continuous.*

Even before Dirichlet proved this theorem, DuBois Reymond gave an example of a continuous f for which $\limsup_{n\to\infty} S_n(f,0) = \infty$. This ruled out the possibility that continuity is sufficient for Equation 1.2 to hold. The difficulty in answering the question affirmatively came in proving convergence of $S_n(f,t)$ as $n \to \infty$. Fejer answered a more relaxed version of the problem, namely, when can f be *reconstructed* from $\hat{f}(r)$ in possibly other ways? He showed that if f satisfies certain conditions even weaker than continuity, then it can be reconstructed from $\hat{f}(r)$ by taking averages.

Definition 1.3 (Cesaro Means). Let $a_1, a_2, \ldots$ be a sequence of real numbers. Their k^{th} *Cesaro mean* is $b_k = (1/k)\sum_{j=1}^{k} a_j$.

Proposition 1.3. *Let $a_1, a_2, \ldots$ be a sequence of real numbers that converges to a. Then the sequence $b_1, b_2, \ldots$ of its Cesaro means converges to a as well. Moreover, the sequence $\{b_i\}$ can converge even when the sequence $\{a_i\}$ does not (e.g. $a_{2j} = 1, a_{2j+1} = 0$).*

Let us apply the idea of Cesaro means to S_n. Define

$$\begin{aligned}
\sigma_n(f,t) &= \frac{1}{n+1}\sum_{k=0}^{n} S_k(f,t) \\
&= \frac{1}{n+1}\sum_{k=0}^{n}\sum_{r=-k}^{k} \hat{f}(r)\, e^{irt} \\
&= \sum_{r=-n}^{n} \frac{n+1-|r|}{n+1}\, \hat{f}(r)\, e^{irt}
\end{aligned}$$

Theorem 1.4 (Fejer). *Let $f : \mathbb{T} \to \mathbb{C}$ be Riemann integrable. If f is continuous at $t \in \mathbb{T}$, then $\lim_{n\to\infty} \sigma_n(f,t) = f(t)$. Further, if f is continuous then the above holds uniformly.*

Proof. Note that $\lim_{n\to\infty} \sigma_n(f,t) = f(t)$ means that $\forall \epsilon > 0, \exists n_0 : n > n_0 \Rightarrow |\sigma_n(f,t) - f(t)| < \epsilon$. The convergence is uniform if the same n_0 works for all t simultaneously. The proof of the Theorem uses Fejer's kernels K_n that behave as continuous approximations to the Dirac delta function.

$$\begin{aligned}
\sigma_n(f,t) &= \sum_{r=-n}^{n} \frac{n+1-|r|}{n+1}\,\hat{f}(r)\,e^{irt} \\
&= \sum_{r=-n}^{n} \frac{n+1-|r|}{n+1}\left(\frac{1}{2\pi}\int_{\mathbb{T}} f(x)\,e^{-irx}\,dx\right)e^{irt} \\
&= \frac{1}{2\pi}\int_{\mathbb{T}} f(x)\sum_{r=-n}^{n}\frac{n+1-|r|}{n+1}\,e^{ir(t-x)}\,dx \\
&= \frac{1}{2\pi}\int_{\mathbb{T}} f(x)K_n(t-x)\,dx \quad \text{for } K_n(z) \stackrel{\text{def}}{=} \sum_{r=-n}^{n}\frac{n+1-|r|}{n+1}\,e^{irz} \\
&= \frac{1}{2\pi}\int_{\mathbb{T}} f(t-y)K_n(y)\,dy \quad \text{for } y = t-x
\end{aligned}$$

which is the convolution of f with kernel K_n. Note that if K_n were the Dirac delta function, then $\int_{\mathbb{T}} f(t-y)K_n(y)\,dy$ would evaluate exactly to $f(t)$. Fejer's kernels K_n approximate this behavior.

Proposition 1.5. *K_n satisfies the following:*

$$K_n(s) = \begin{cases} \frac{1}{n+1}\left(\frac{\sin\frac{n+1}{2}s}{\sin\frac{n}{2}}\right)^2 & \text{if } s \neq 0 \\ n+1 & \text{if } s = 0 \end{cases}$$

The kernels K_n have three useful properties.

1. $\forall u : K_n(u) \geq 0$
2. $\forall \delta > 0 : K_n(s) \to 0$ uniformly outside the interval $[-\delta,\delta]$, i.e. $\forall \epsilon > 0, \exists n_0 : s \notin [-\delta,\delta] \Rightarrow |K_n(s)| < \epsilon$
3. $(1/2\pi)\int_{\mathbb{T}} K_n(s)\,ds = 1$

Given any $\epsilon > 0$, we seek a large enough n_0 such that for all $n > n_0$, $|\int_{\mathbb{T}} f(t-y)K_n(y)\,dy - f(t)| < \epsilon$. Divide this integral into two intervals:

$$\int_{\mathbb{T}} f(t-y)K_n(y)\,dy = \int_{-\delta}^{\delta} f(t-y)K_n(y)\,dy + \int_{\mathbb{T}\setminus[-\delta,\delta]} f(t-y)K_n(y)\,dy$$

The first integral on the RHS converges to $2\pi f(t)$ because $f(t-y)$ is almost constant and equals $f(t)$ in the range $y \in [-\delta,\delta]$ and $\int_{\mathbb{T}\setminus[-\delta,\delta]} K_n(s)\,ds$ converges to 2π because of property 3 of K_n. The second integral converges to 0 because f is bounded and because of property 2 of K_n. Hence the LHS converges to $2\pi f(t)$, finishing the proof. □

Corollary 1.6. *If $f, g : \mathbb{T} \to \mathbb{C}$ are continuous functions and $\forall r \in \mathbb{Z} : \hat{f}(r) = \hat{g}(r)$, then $f = g$.*

Proof. Let $h \stackrel{\text{def}}{=} f - g$. h is also continuous. $\forall r : \hat{h}(r) = \hat{f}(r) - \hat{g}(r) = 0$. By Fejer's Theorem, $h \equiv 0$. □

1.4.3 Connection with Weierstrass' Theorem

Because of the uniform convergence part of Fejer's Theorem, we have proved that for all $f : \mathbb{T} \to \mathbb{C}$ continuous and all $\epsilon > 0$, there exists a trigonometric polynomial P such that for all $t \in \mathbb{T}$, $|f(t) - P(t)| < \epsilon$. This implies **Weierstrass' Theorem** which states that "under $l_\infty[a, b]$ norm, polynomials are dense in $C[a, b]$," i.e., for all $f : [a, b] \to \mathbb{R}$ continuous and all $\epsilon > 0$, there exists a polynomial P such that for all $x \in [a, b]$, $|f(x) - P(x)| < \epsilon$.

Informally, Weierstrass' Theorem says that given any continuous function over a finite inverval and an arbitrarily small envelope around it, we can find a polynomial that fits inside that envelope in that interval. To see why this is implied by Fejer's Theorem, simply convert the given function $f : [a, b] \to \mathbb{C}$ into a symmetric function g over an interval of size $2(b - a)$, identify the end points of the new interval so that it is isomorphic to $\mathbb{T}$, and use Fejer's Theorem to conclude that $\sigma_n(g, .)$ is a trigonometric polynomial close to g (and hence f). To see why Weierstrass' Theorem implies Fejer's Theorem, recall that $\cos rt$ can be expressed as a degree r polynomial in $\cos t$. Use this to express the promised trigonometric polynomial $P(t)$ as a linear combination of $\cos rt$ and $\sin rt$ with $-n \leq r \leq n$.

Remark. Weierstrass' Theorem can alternatively be proved using Bernstein's polynomials even though normal interpolation polynomials do not work well for this purpose. Consider $f : [0, 1] \to \mathbb{R}$. The n^{th} Bernstein polynomial is $B_n(f, x) \stackrel{\text{def}}{=} \sum_{k=0}^{n} f(\frac{k}{n})\binom{n}{k}x^k(1 - x)^{n-k}$. The key idea is that this involves the fact that the Binomial distribution $P(k) = \binom{n}{k}x^k(1-x)^{n-k}$ is highly concentrated around $k = xn$ and thus approximates the behavior of the Dirac delta function.