Lecture 3

Harmonic Analysis on the Cube and Parseval's Identity

Jan 28, 2005
Lecturer: Nati Linial
Notes: Pete Couperus and Neva Cherniavsky

3.1 Where we can use this

During the past weeks, we developed the general machinery which we will apply to problems in discrete math and computer science in the following weeks. In the general setting, we can ask how much information can we determine about a function f given its Fourier coefficients $\hat{f}$. Or, given f what can we say about $\hat{f}$? There is some distinction between properties which will hold in the general setting, and those that make sense for the specific spaces we have dealt with. So far, we have looked at

1. $\mathbb{T}$ (the unit circle/Fourier Series).
2. $\mathbb{Z}/n\mathbb{Z}$ (Discrete Fourier Transform).
3. $\mathbb{R}$ (Real Fourier Transform).
4. $\{0,1\}^n = \mathbf{GF}(2)^n = (\mathbb{Z}/2\mathbb{Z})^n$ (the n-cube).

For the n-cube (or for any space we wish to do Harmonic Analysis on), we need to determine the characters. We can view elements of $\{0,1\}^n$ as subsets of $[n] = \{1, ..., n\}$, and then to each subset $S \subseteq [n]$, let $\chi_S(T) = (-1)^{|S \cap T|}$. Then:

$$\langle \chi_{S_1}, \chi_{S_2} \rangle = \frac{1}{2^n} \sum_{T \subseteq [n]} (-1)^{|S_1 \cap T| + |S_2 \cap T|}$$

To see that the χ_S form an orthonormal basis, suppose that $x \in S_1 - S_2$. Then, the function

$$\phi(A) = \begin{cases} A - \{x\} & x \in A \\ A \cup \{x\} & x \notin A \end{cases}$$

gives a bijection between $\{A : |S_1 \cap A| \equiv |S_2 \cap A| \pmod 2\}$ and $\{A : |S_1 \cap A| \not\equiv |S_2 \cap A| \pmod 2\}$. So, $\langle \chi_{S_1}, \chi_{S_2} \rangle = 0$ for $S_1 \neq S_2$. If $S_1 = S_2$, then $|T \cap S_1| + |T \cap S_2|$ is always even, so $\langle \chi_S, \chi_S \rangle = 1$.

Hence, the χ_S form an orthonormal basis for functions from $\{0,1\}^n \to \mathbb{R}$. (This is, of course, true in general, but it's useful to see this explicitly for this special case). Then for any $f : \{0,1\}^n \to \mathbb{R}$, we can write $f = \sum_{S \subseteq [n]} \hat{f}(S)\chi_S$, where

$$\hat{f}(S) = \langle f, \chi_S \rangle = \frac{1}{2^n} \sum_{T \subseteq [n]} f(T) \cdot (-1)^{|S \cap T|}.$$

There is an equivalent and often useful way of viewing this. We can also view the n-cube as $\{-1,1\}^n$ with coordinate-wise multiplication. In this case, any function $f : \{-1,1\}^n \to \mathbb{R}$ can be uniquely expressed as a multilinear polynomial:

$$f = \sum_{S \subset \{0,1\}^n} a_S \prod_{i \in S} x_i$$

where $\prod_{i \in S} x_i$ corresponds to χ_S.

There is an advantage to the fact that we now deal with a finite group. Note that $f = \sum_{S \subseteq [n]} \hat{f}(S)\chi_S$ is always the case for functions over the n-cube, unlike working over $\mathbb{T}$. Working over $\mathbb{T}$, we made some assumptions on f to be able have a similar formula to recover f from its fourier coefficients.

Now we can ask, what can be said about $\hat{f}$ when f is boolean (when the range of f is $\{0,1\}$)? More specifically, how do the properties of f get reflected in $\hat{f}$? In general, this is too hard a question to tackle. But what sorts of relationships between properties are we looking for? In the case of $\mathbb{T}$, the smoothness of f roughly corresponds to its fourier coefficients $\hat{f}(r)$ decaying rapidly as $r \to \infty$. E.g.

$$\begin{array}{rcl} f : \mathbb{T} \to \mathbb{C} & \leftrightarrow & \{f(r) | r \in \mathbb{Z}\} \\ \text{smooth} & \leftrightarrow & \hat{f}(r) \text{ decays rapidly} \end{array}$$

An instance of this relationship can be seen from the following theorems.

Theorem 3.1. *Let $f : \mathbb{T} \to \mathbb{C}$ be continuous, and suppose that $\sum_{r=-\infty}^{\infty} |\hat{f}(r)|$ converges. Then $S_n(f) \to f$ uniformly.*

We can derive this theorem from another.

Theorem 3.2. *Suppose that the sequence $\sum_{r=-n}^{n} |a_r|$ converges (as $n \to \infty$). Then $g_n(t) = \sum_{r=-n}^{n} a_r e^{irt}$ converges uniformly as $n \to \infty$ on $\mathbb{T}$ to $g : \mathbb{T} \to \mathbb{C}$, where g is continuous and $\hat{g}(r) = a_r$ for all r.*

This (roughly) says that if we have a sequence that is decreasing rapidly enough (its series converges absolutely), then we can choose these to be the Fourier coefficients for some continuous function.

To see that Theorem 3.2 implies Theorem 3.1, if $\hat{f}(r) = \hat{g}(r) = a_r$ for all r, and both f and g are continuous, then $f = g$. This is based on Fejer's Theorem (or Weierstrauss).

So to prove Theorem 3.1, all that remains is to prove Theorem 3.2.

Proof. The underlying idea for the proof of Theorem 3.2 is that $C(\mathbb{T})$ with the ∞-norm is a **complete metric space**, meaning that all Cauchy sequences converge. Recall, a sequence (a_n) is Cauchy if for $\epsilon > 0$, there is some N so for $n, m \geq N$, we have $d(a_n, a_m) < \epsilon$ (where d is whatever metric we are using). So, to prove the theorem, we only need to check that $\{f_n\} = \{\sum_{-n}^{n} a_r e^{irt}\}$ is a Cauchy sequence with the ∞-norm. Since $s_n := \sum_{r=-n}^{n} |a_r|$ converges, for $\epsilon > 0$, there is some N so that $|s_m - s_n| < \epsilon$ for $n, m \geq N$ (basically, the tail end is small), hence

$$\left\| \sum_{m \geq |r| > n} a_r e^{irt} \right\| \leq \left| \sum_{m \geq |r| > n} |a_r| \right| < \epsilon.$$

So, the $\{g_n\}$ forms a Cauchy sequence.

Hence,

$$\begin{aligned} \sum_{r=-n}^{n} a_r e^{-irt} &\to g \text{ uniformly, so} \\ e^{-ikt} \sum_{r=-n}^{n} a_r e^{-irt} &\to e^{-ikt} g(t) \text{ uniformly.} \\ \int_{\mathbb{T}} e^{-ikt} \sum_{r=-n}^{n} a_r e^{-irt} dt &\to \int_{\mathbb{T}} e^{-ikt} g(t). \end{aligned} \tag{3.1}$$

□

Recall, du Bois Raymond gives an example of $f : \mathbb{T} \to \mathbb{C}$ such that $\overline{\lim} |S_n(f, 0)| = +\infty$. However, if the first derivative is somewhat controlled, we can say more.

Theorem 3.3. *Let $f : \mathbb{T} \to \mathbb{C}$ be continuous and suppose that f' is defined for all but a finite subset of $\mathbb{T}$. Then $S_n(f) \to f$ uniformly.*

f smooth $\leftrightarrow$ $\hat{f}$ decays rapidly $\Rightarrow$ "$S_n f \to f$".

Recall from basic analysis, if f_n are continuously differentiable and if $f_n \to f$ uniformly and $f_n' \to g$ uniformly then $f' = g$ and g is continuous. This will allow us to show that the Fourier Series of f' is attained by termwise derivatives of the Fourier Series of f.

Theorem 3.4. *Let $f : \mathbb{T} \to \mathbb{C}$ be continuous and suppose that $\sum_{r=-\infty}^{\infty} r|\hat{f}(r)|$ converges. Then f is continuously differentiable and $\sum_{r=-n}^{n} ir\hat{f}(r)e^{irt} \to f'$ uniformly.*

Proof. We would like to show that we can apply this when $f_n = S_n f$. But if $\sum_{r=-\infty}^{\infty} r|\hat{f}(r)|$ converges, then $\sum_{r=-\infty}^{\infty} |\hat{f}(r)|$ converges (since the each term is smaller). So we have

$$|\hat{f}(r)| \leq |r\hat{f}(r)| \Rightarrow \sum_{-n}^{n} |\hat{f}(r)| \text{ converges}$$

So by Theorem 3.1, $f_n = S_n f \to f$ uniformly. By the same theorem, $f'_n = \sum_{r=-n}^{n} ir\hat{f}(r)e^{irt} \to g$ is continuous. By the statement above due to basic analysis, we know that this implies that f is continuously differentiable. $\square$

A similar argument will provide a stronger connection between the idea that $\hat{f}(r)$ are rapidly decreasing implies that f is "smoother".

Proposition 3.5. *Let $f : \mathbb{T} \to \mathbb{C}$ satisfy $f^{(n-1)}$ is continuously differentiable except possibly finitely many points X, and $|f^{(n)}(x)| \leq M$ for $x \notin X$. Then $\forall r \neq 0 |\hat{f}(r)| \leq Mr^{-n}$.*

Proof. (Integration by parts).

$$\hat{f}(r) = \frac{1}{2\pi}\int_{\mathbb{T}} f(t)e^{-irt}dt. \text{ Let } u = f(t), dv = e^{-irt}dt. \text{ Then } du = f'(t)dt, v = \frac{e^{-irt}}{-ir}.$$

$$\begin{aligned}\hat{f}(r) = \frac{1}{2\pi}\int_{\mathbb{T}} f(t)e^{-irt}dt = \\ \frac{1}{2\pi}\left[f(t)\frac{e^{-irt}}{-ir}\Big|_{-\pi}^{\pi} - \int_{\mathbb{T}} f'(t)\frac{e^{-irt}}{-ir}dt\right] = \\ \frac{1}{2\pi}\left[0 - \int_{\mathbb{T}} f'(t)\frac{e^{-irt}}{-ir}dt\right] = \cdots = \text{ (first term is 0 since } f \text{ is periodic)} \\ \frac{1}{2\pi(-ir)^n}\int_{\mathbb{T}} f^{(n)}(t)e^{-irt}dt.\end{aligned} \tag{3.2}$$

So

$$\left|\hat{f}(r)\right| \leq \left|\frac{(-ir)^{-n}}{2\pi}\right| \int_{-\pi}^{\pi} |f^{(n)}(t)e^{-irt}|dt = O(\frac{1}{r^n})$$

$\square$

Corollary 3.6. *If $f : \mathbb{T} \to \mathbb{C}$ is in C^2 (twice continuously differentiable), then $S_n f \to f$ uniformly.*

Proof.

$$\hat{f}(r) = O\left(\frac{1}{r^2}\right) \Rightarrow \sum_{r=-\infty}^{\infty}\left|\hat{f}(r)\right| \text{ converges.}$$

So, $S_n f \to f$ uniformly. $\square$

3.2 Rate of Convergence

Until now, we haven't really addressed the rate of convergence, meaning when $S_n(f)$ does converge to f, how fast does it converge to f? Examine $g(x) = \pi - |x|$ for $x \in [-\pi, \pi]$, and extend g periodically to $h(x)$. Direction calculation gives $|S_n(h, 0) - \pi| > \frac{1}{n+2}$. By using L_2 theory, it can be further shown that every trigonometric polynomial P of degree n has the property $\|P - h\|_\infty > \Omega(n^{-3/2})$. Kolmogorov showed the following.

Theorem 3.7. *(Kolmogorov) For all $A > 0$, there is a trigonometric polynomial f such that:*

1. $f \geq 0$.
2. $\frac{1}{2\pi} \int_{\mathbb{T}} f(t)dt \leq 1$
3. *For every* $x \in \mathbb{T}$, $\sup_n |S_n(f,x)| \geq A$.

Hence, there is a Lebesgue integrable function f such that for all $x \in \mathbb{T}$, $\overline{\lim} |S_n(f,x)| = +\infty$.

3.2.1 Convergence Results

In 1964, Carleson proved the following.

Theorem 3.8. *(Carleson) If f is continuous (or only Riemann integrable), then $S_n f \to f$ almost everywhere.*

Later, Kahane and Katznelson proved that this result is tight.

Theorem 3.9. *For all $E \subseteq \mathbb{T}$ with $\mu(E) = 0$, there is a continuous f such that $S_n f \to f$ exactly on $\mathbb{T} - E$.*

Notice that these results make somewhat weak assumptions on f. We will now work on seeing how things improve in the situation where f is an L_2 function.

3.3 L_2 theory for Fourier Series

Recall part of original question was "how are f and $\hat{f}$ related"? Our immediate goal will be to show that in the L_2 case, their norms are identical, which is the Parseval identity. Recall, $\|f\|_2 = \sqrt{\int_{\mathbb{T}} |f(t)|^2 dt}$. Then the Parseval identity states $\|f\|_2 = \|\hat{f}\|_2$. For the Discrete Fourier Transform, this essentially means that the transform matrix is an orthonormal matrix.

We will procede by focusing on Hilbert Spaces. A Hilbert Space $\mathcal{H}$ is a normed ($\mathbb{C}$-linear) space with an inner product $\langle \cdot, \cdot \rangle$ satisfying the following axioms.

1. $\langle ax + by, z \rangle = a\langle x, z \rangle + b\langle y, z \rangle$.
2. $\langle x, y \rangle = \overline{\langle y, x \rangle}$.
3. $\langle x, x \rangle = \|x\|^2 \geq 0$ with equality $\iff x = 0$.

There are a number of facts that we know about familiar Hilbert spaces (like $\mathbb{R}^n$) that hold for general Hilbert spaces as well.

Theorem 3.10. *If $\mathcal{H}$ is a Hilbert space, then the Cauchy-Schwarz Inequality holds, namely if $f, g \in \mathcal{H}$, then $\|f\| \cdot \|g\| \geq \langle f, g\rangle$.*

Proof. We will show the proof for real Hilbert spaces.

$$0 \leq \langle f-\lambda g, f-\lambda g\rangle = \|f\|^2 - 2\lambda\langle f, g\rangle + \lambda^2\|g\|^2. \tag{3.3}$$

Viewing this as a degree 2 polynomial in λ, it is non-negative, so has at most one real root. Hence, the discriminant $(-2\langle f, g\rangle)^2 - 4\|f\|^2\|g\|^2 \leq 0$. Hence, $\langle f, g\rangle^2 \leq \|f\|^2\|g\|^2$. □

One may ask, if we have an element $f \in \mathcal{H}$, how can we best approximate f with respect to some basis? Specifically, let $e_{-n}, ..., e_0, e_1, ..., e_n$ be an orthonormal system in $\mathcal{H}$ (meaning, $\langle e_i, e_j\rangle = \delta_{i,j}$). Given $f \in \mathcal{H}$, the question is to find $\lambda_i \in \mathbb{C}$ such that $\|f - \sum_i \lambda_i e_i\|$ is minimized.

Theorem 3.11. *Let $\mathcal{H}$, $\{e_i\}$, f be as above. Set $g = \sum_{j=-n}^n \lambda_j e_j$, and $g_0 = \sum_{j=-n}^n \langle f, e_j\rangle e_j$. Then*

$$\|f\|_2^2 \geq \sum_{j=-n}^{n} \langle f, e_j\rangle^2, \|f-g\|_2 \geq \|f-g_0\|_2 = \sqrt{\|f\|_2^2 - \sum_{j=-n}^{n}\langle f, e_j\rangle^2} \tag{3.4}$$

with equality iff $\lambda_j = \langle f, e_j\rangle$ for all j.

Proof.

$$\begin{aligned}\|f-g\|_2^2 = \langle f-g, f-g\rangle = \langle f - \sum_j \lambda_j e_j, f - \sum_j \lambda_j e_j\rangle = \\ \|f\|_2^2 - (\sum_j \lambda_j \overline{\langle f, e_j\rangle} + \overline{\lambda_j}\langle f, e_j\rangle) + \sum_j |\lambda_j|^2 = \\ \langle f, f\rangle + \sum_j |\lambda_j - \langle f, e_j\rangle|^2 - \sum_j |\langle f, e_j\rangle|^2 \geq \\ \langle f, f\rangle - \sum_j |\langle f, e_j\rangle|^2 = \|f - g_0\|^2.\end{aligned} \tag{3.5}$$

Note that equality in the last step occurs exactly when $\lambda_j = \langle f, e_j\rangle$ for all j. □

Corollary 3.12. *(Approximation and Bessel's Inequality).*

1. *$S_n f$ is the closest (in the L_2 sense) degree n trigonometric polynomial approximation to f.*
2. *(Bessel's Inequality). If $f \in L_2(\mathbb{T})$, then*

$$\|f\|^2 = \frac{1}{2\pi}\int_{\mathbb{T}} |f(t)|^2 dt \geq \sum_{r=-n}^{n} |\hat{f}(r)|^2,$$

and $\|f\|^2 \geq \sum_{r=-\infty}^{\infty} |\hat{f}(r)|^2$.

This shows one side of the Parseval Identity, namely $\|f\|^2 \geq \|\hat{f}\|^2$.
Recall by Theorem 3.1, if f continuous and $\hat{f} \in l_1$ (meaning $\sum_r |\hat{f}(r)|$ converges), then $S_n f \to f$ uniformly. We will show that f having continuous first derivative in fact implies that $\hat{f}$ is in l_1.

Corollary 3.13. *If $f \in C^1$, then $S_n f \to f$ uniformly.*

Proof.

$$\begin{aligned}
\sum_{r=-n}^{n} |\hat{f}(r)| &= |\hat{f}(0)| + \sum_{1\leq|r|\leq n} |r\hat{f}(r)| \cdot \frac{1}{|r|} \\
\text{(by Cauchy-Schwartz)} \quad &\leq |\hat{f}(0)| + \sqrt{2\sum_{1\leq r\leq n} \frac{1}{r^2}} \cdot \sqrt{\sum_{1\leq|r|\leq n} |\hat{f'}(r)|^2} \\
\text{(by identity } \sum_{1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}) \quad &\leq |\hat{f}(0)| + \sqrt{\frac{\pi^2}{3} \cdot \frac{1}{2\pi} \int_{\mathbb{T}} |f'(t)|^2 dt}
\end{aligned}$$

which is bounded since the first derivative is bounded (continuous on $\mathbb{T}$). □

3.3.1 Parseval's Identity

We are now ready to complete the proof of the Parseval Identity.

Theorem 3.14. *If $f : \mathbb{T} \to \mathbb{C}$ is continuous, then $\|f - S_n f\| \to 0$.*

Proof. By Weierstrass (or Fejer) approximation, for any $\epsilon > 0$, there is some trigonometric polynomial P such that $\|f - P\|_\infty < \epsilon$. So,

$$\|f - S_n f\|_2 \leq \|f - P\|_2 + \|S_n P - S_n f\|_2 \leq \|f - P\|_\infty + \|S_n(P - f)\|_2$$

We use the fact that $S_n P = f$ for every trigonometric polynomial of degree $\leq n$. Then Bessel's inequality tells us $\|S_n(P-f)\|_2 \leq \|P-f\|_2$. Since $\|P-f\|_2 \leq \|P-f\|_\infty < \epsilon$, we see that $\|f - S_n f\|_2 < 2\epsilon$. This completes the proof. □

Hence, it is easy to see $\|f\|_2 = \|\hat{f}\|_2$, and we have the Parseval Identity.

Corollary 3.15. *(Parseval) If $f : \mathbb{T} \to \mathbb{C}$ is continuous, then $\frac{1}{2\pi}\int_{\mathbb{T}} |f(t)|^2 dt = \|f\|_2^2 = \sum_{r=-\infty}^{\infty} |\hat{f}(r)|^2$.*

Proof. Since $\|f - S_n f\|_2^2 = \|f\|_2^2 - \sum_{r=-n}^{n} |\hat{f}(r)|^2$ goes to 0 as $n \to \infty$, we conclude that $\sum_{r=-n}^{n} |\hat{f}(r)|^2 \to \|f\|_2^2$ as $n \to \infty$. □

In other words, $f \to \hat{f}$ is an isometry in L_2.

3.4 Geometric Proof of the Isoperimetric Inequality

We will complete this next time. Here, we will present Steiner's idea to resolve the following question. What is the largest area of a planar region with fixed circumference L? We will present Steiner's idea here. Suppose that C is a curve such that the area enclosed is optimal.

C **encloses a convex region.**

If not, then there are points A, B such that the line segment joining A and B lies outside the region. By

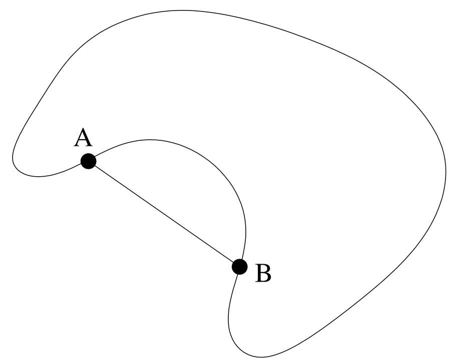

Figure 3.1: Joining A and B yields more area.

replacing the arc from A to B with the line segment from A to B, we increase the area, and decrease the circumference. See Figure 3.1.

C **encloses a centrally symmetric region.**

If not, pick points A, B such that the arc length from A to B is the same for both directions travelled.

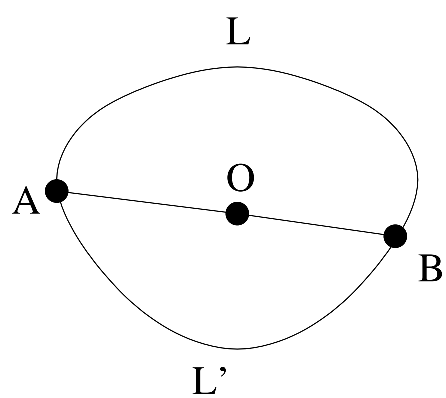

Figure 3.2: Reflecting larger area yields more area.

Suppose that the region enclosed by $\overline{AB} \cup L$ has area at least that of the region enclosed by $\overline{AB} \cup L'$. We can then replace the latter by a mirror copy of the first. This can only increase the total area and yields a region that is centrally symmetric with respect to the middle of the segment $[A, B]$. C **is a circle.**

Recall the following fact from Euclidean geometry: A circle is the locus of all points x such that xA is perpendicular to xB where AB is some segment (which is the diameter of the circle). Therefore, if this is not so, then there is some parallelogram a, b, c, d, with $\overline{ac}$ passing the the center, inscribed in C (since C is centrally symmetric), and with the angle at b not equal to $\frac{\pi}{2}$. Now, "move" sides a, b and c, d to sides a', b'

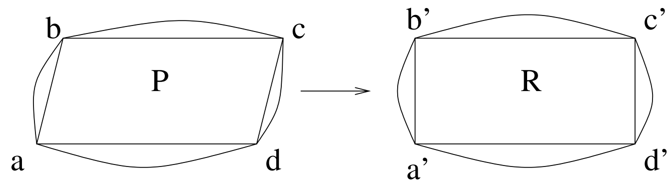

Figure 3.3: Changing Parallelogram to Rectangle Yields more area.

and c', d' such that a', b', c', d' forms a rectangle. See Figure 3.3. We obtain a new curve C' such that the area outside of rectangle $R = [a', b', c', d']$ is the same as the area outside of the parallelogram $P = [a, b, c, d]$. Since the side lengths of R and P are the same, the area enclosed by R must exceed the area enclosed by P, so the area enclosed by C' must exceed the area enclosed by C. Hence, C was not optimal. Hence, our parallelogram P must have angles equal to $\frac{\pi}{2}$.

Although these ideas are pretty and useful, this is still not a proof of the isoperimetric inequality. We do not know that an optimal C exists, only that if it does, it must be a circle.