ALMOST-EVERYWHERE CONVERGENCE FOR ALMOST EVERYONE

JOEL H. SHAPIRO

OCTOBER 16, 2022

In 1926 Stefan Banach [3] set out a "two-step program" for proving theorems about almost-everywhere convergence. Banach's program lies at the heart of all modern work on a.e. convergence. In these notes we'll see why.

Contents

1 Introduction

Almost a century ago, Stefan Banach published a paper [3] in which he showed how to approach the notion of "almost-everywhere convergence" through the emerging field of functional analysis. Banach considered sequences (T_n) of measure-continuous[1] linear transformations from a Banach space B into the space M of (a.e. equivalence classes of) real-valued Lebesgue-measurable functions on the unit interval, and asked when $(T_n f)$ could be expected to converge a.e. for every $f \in B$.

Throughout these notes we use "a.e. convergence " to mean "a.e. convergence to a finite value."

[1] For a map $T\colon B \to M$ to be "measure-continuous" (or "continuous in measure") means that $Tf_n \to Tf$ in measure whenever $f_n \to f$ in B.

Banach's paper establishes the dependence of this problem on a favorable estimate for the *maximal operator* $T^\#$, defined by:

$$T^\# f := \sup_n |T_n f| \qquad (f \in B). \tag{1}$$

He proves three main theorems; the first, a *necessary condition* for a.e. convergence, reveals the sort of "favorable estimate" that will dominate the rest of the discussion.

1.1 Theorem.[2] *Suppose $(T_n f)$ converges a.e. for each $f \in B$. Then*

$$\lim_{\lambda\to\infty} \sup_{f\in B} \mu\{T^\# f > \lambda \|f\|\} = 0, \tag{2}$$

where μ denotes Lebesgue measure on $[0, 1]$.

[2] Banach [3], Théoréme I, pp. 356-7.

Throughout these notes we use the notation $\{g > \lambda\}$ to denote $\{x \in X\colon g(x) > \lambda\}$.

To get some feeling for condition (2), define the *distribution* of a non-negative, measurable function f on $[0, 1]$ to be the function $\gamma_f\colon [0, \infty] \to [0, 1]$ given by

$$\gamma_f(\lambda) = \mu\{f > \lambda\} \qquad (\lambda \geq 0). \tag{3}$$

Later on we'll see that certain important integrals involving f can be rewritten as Riemann integrals involving its distribution. For example, if $1 \leq p < \infty$, then

$$\int_0^1 f(x)^p\, dx = p \int_0^\infty \lambda^{p-1} \gamma_f(\lambda)\, d\lambda$$

(see, e.g., Corollary **??**, page **??**).

Clearly γ_f decreases monotonically on $[0, \infty]$, and it's easy to see that

$$f < \infty \text{ a.e.} \iff \lim_{\lambda\to\infty} \gamma_f(\lambda) = 0. \tag{4}$$

The backward implication of (4) holds for all measures μ, the forward one only for finite measures.

Theorem 1.1 seems to be a sort of "uniform-boundedness result" in that asserts that a.e. convergence of $(T_n f)$ for each $f \in B$ (i.e., "pointwise a.e. convergence") implies that the distribution function of $T^\# f$ converges to 0 *uniformly for f in the unit ball of B.* It should therefore come as no surprise that the proof we give here (Theorem 5.1, page 15) will involve the Baire Category Theorem.

Banach's original proof used a more complicated "gliding hump" argument to show that if (2) were to fail, then $(T_n f)$ would not converge a.e. for some $f \in B$.

1.2 Theorem.[3] *If $(T_n f)$ converges a.e. for each $f \in B$, then the "limit operator" $T\colon B \to M$ defined by*

$$Tf(x) = \lim_n T_n f(x) \qquad (f \in B,\ x \in [0, 1]) \tag{5}$$

is measure-continuous.

[3] Banach [3], Théoréme II page 359.

Sketch of proof. Equation (5) along with Theorem 1.1 implies the maximal condition (2), which quickly implies that $T^{\#}$ is measure-continuous, and this property transfers easily to the limit operator T defined by (5).

1.3 Theorem.[4] *Suppose:*

[4] Banach [3], Théoréme III page 359.

(a) $T^{\#} f < \infty$ *a.e. for each* $f \in B$*, and*

(b $\lim_n T_n f$ *exists a.e. for each* f *in a dense subset of* B*. Then that limit exists a.e. for every* $f \in B$*.*

This, the most important result result in [3], gives us a program for proving a.e. convergence:

First: Prove that $\lim_n T_n f$ exists a.e. for a dense subset of vectors $f \in B$ (usually this is the "easy" part).

Next: Prove that the maximal function $T^{\#}$ obeys the maximal condition (2).

Then: We're done! The uniformity baked into (2) allows the desired a.e. convergence to be extended from the dense subset of B to the whole space.

We'll discuss "Banach's program" for proving a.e. convergence (a.k.a. Theorems 1.2 and 1.3) carefully in the next section, after which we'll show how it suffuses three famous a.e. convergence theorems:

Section 3: Lebesgue's theorem on "differentiation of integrals."

Section 6: Boundary convergence in the Dirichlet Problem (for the upper half plane).

Section 4: Birkhoff's "Pointwise Ergodic Theorem."

We'll give the "Baire Category" proof of Banach's necessary condition for a.e. convergence (Theorem 1.1) in §5, along with discussion of later refinements. Finally, we'll close with a section devoted to L^p estimates for the previously-appearing maximal functions.

It will turn out that all of them belong to L^∞, but none to L^1. The question arises: What about L^p for $1 < p < \infty$?

Part I: Basics

2 *Banach's Program: Sufficiency*

Although Banach worked only with Lebesgue measure on $[0,1]$, his apply to very general measure spaces. Thus our interpretation of his work in [3] will involve:

(a) A Banach space B, with norm $\|\cdot\|$.

Throughout these notes, B will be a *real* Banach space. The passage to complex scalars is easily achieved by applying the "real" results to real and imaginary parts.

(b) A measure space $(X, \mathscr{F}, \mu)$, and its associated space M of (a.e. equivalence classes of) scalar-valued measurable functions.

(c) A sequence $(T_n)_1^\infty$ of *sublinear*[5] transformations $B \to M$, each of which is "continuous in measure" (as described in sidenote 1 on page 2), for which we define the "maximal operator" $T^\#$ by:

(1) $$T^\# f := \sup_n |T_n f| \qquad (f \in B),$$

Note that $T^\#$ takes B into the collection of non-negative functions on X that are $\mathscr{F}$-measurable, but possibly not a.e. finite-valued.

[5] To say a mapping $T\colon B \to M$ is *sublinear* means that whenever $f, g \in M$ and $a \in \mathbb{R}$:

- $|T(f+g)| \le |Tf| + |Tg|$ a.e., and
- $T(af) = |a|\,|T(f)|$ a.e.

For example, the maximal function $T^\#$ defined by (1) is sublinear, as is the map $f \to |Tf|$ for any linear $T\colon B \to M$.

2.1 Theorem.[6] *Suppose that:*

[6] Banach [3], Theoréme II and Theoréme III, page 359. Here: Theorems 1.2 and 1.3.

(a) $\lim_n T_n f$ *exists (finitely) a.e. for each f in a dense subset D of B.*

and

(b) *The maximal function $T^\#$ obeys the "uniform maximal condition"*

(2) $$\lim_{\lambda\to 0} \sup_{f\in B} \mu\{T^\# f > \lambda \|f\|\} = 0\,.$$

Then $\lim_n T_n f$ *exists (finitely) for each $f \in B$, and the limit operator T defined by*

(5′) $$Tf = \lim_{n\to\infty} T_n f \qquad (f \in B,\ x \in X)$$

is a measure-continuous, sublinear map that taking B into M.

Proof. To begin the proof of part (a) let's define, for $f \in B$:

(6) $$\Delta f(x) := \limsup_n T_n f(x) - \liminf_n T_n f(x) \qquad (x \in X)\,,$$

and for $\lambda > 0$

(7) $$\beta(\lambda) := \sup_{f\in B} \mu\{T^\# f > \lambda \|f\|\}$$

Observe that the right-hand side of (6) does not have an "$\infty - \infty$" problem because, $T^\# f < \infty$ a.e. thanks to inequality (2), hence both terms in question are a.e. finite.

Thus we are assuming that $\lim_{\lambda\to\infty} \beta(\lambda) = 0$, and wish to show that $\Delta f = 0$ a.e. for each $f \in B$. This is already true for each $f \in D$, so we're faced with the familiar problem of using a "uniform condition" to extend a known result from a dense subset to the whole space.

To this end, note that:

(Δ_a) $0 \le \Delta f \le 2\,T^{\#} f$ a.e. for each $f \in B$.

(Δ_b) $\Delta(f - g) = \Delta f$ a.e. for each $g \in D$ and $f \in B$.

Now fix $f \in B, \lambda > 0$, and $\delta > 0$. Since D is dense in B there exists $g \in D$ such that $\|f - g\| < \delta$. Thus:

$$\begin{aligned} \mu\{\Delta f > \lambda\} &= \mu\{\Delta(f - g) > \lambda\} \\ &\le \mu(\{T^{\#}(f - g) > \lambda\}) \\ &\le \beta\Big(\frac{\lambda}{\|f - g\|}\Big) \\ &\le \beta(\lambda/\delta)\,. \end{aligned}$$

By (Δ_b) above.

By (Δ_a) above.

By the maximal inequality (2)

β is monotone decreasing.

In summary: for each positive λ and δ:

$$\mu\{\Delta f > \lambda\} \le \beta(\lambda/\delta).$$

Upon keeping λ fixed and sending δ to 0, this inequality yields:

$$\mu\{\Delta f > \lambda\} = 0 \quad \text{for each} \quad \lambda > 0.$$

Since any countable union of sets of measure zero has measure zero:

$$\mu\{\Delta f = 0\} = \mu\Big(\bigcup_{k \in \mathbb{N}} \{\Delta f > 1/k\}\Big) = 0,$$

i.e., $\Delta f = 0$ a.e., as we wished to show.

It remains to prove part (b), the measure-continuity of the limit operator T. This will follow from:

(*) *$T^{\#}$ takes B into M, and is measure-continuous.*

Proof of (*). Note that by linearity it's enough to show that if $f_n \to 0$ in B then $T^{\#} f_n \to 0$ in measure. To this end, fix $\lambda > 0$ and use the maximal condition (2):

$$\mu\{T^{\#} f_n > \lambda\} = \mu\Big\{T^{\#} f_n > \frac{\lambda}{\|f_n\|}\|f_n\|\Big\} \le \beta\Big(\frac{\lambda}{\|f_n\|}\Big)$$

Here we use the fact that β is a monotonically decreasing function on $(0, \infty)$.

Since $\lim_n \|f_n\| = 0$ and $\lim_{t\to\infty} \beta(t) = 0$, we have $\lim_n \beta(\lambda/\|f_n\|) = 0$. Thus $\lim_n \mu\{T^{\#} f_n > \lambda\} = 0$ for] $\lambda > 0$, i.e., $T^{\#} f_n \to 0$ in measure as $n \to \infty$. This completes the proof of (*).

To complete the proof that T is measure-continuous, note that sublinearity allows us once again to restrict attention to continuity at the origin. For this, suppose once again that $\lim_n \|f_n\| = 0$. Then for $\lambda > 0$ and n a non-negative integer:

$$\mu\{T f_n > \lambda\} = \mu\{\lim_k T_k f_n > \lambda\} \le \mu\{T^{\#} f_n > \lambda\}.$$

Thus $\lim_n \mu\{T f_n > \lambda\} = 0$, hence $T f_n \to 0$ in measure. This completes the proof of Theorem 2.1. □

3 *The Lebesgue Differentiation Theorem*

The Fundamental Theorem of Integral Calculus tells us that if $f\colon \mathbb{R} \to \mathbb{R}$ is continuous, with compact support, then

$$\frac{d}{dx} \int_{-\infty}^{x} f(t)\, dt = f(x)$$

for each $x \in \mathbb{R}$, i.e.

$$\lim_{r\to 0+} \frac{1}{r} \int_{x}^{x+r} f(t)\, dt = f(x) = \lim_{r\to 0+} \frac{1}{r} \int_{x-r}^{x} f(t)\, dt\,.$$

The argument stems from the uniform continuity of the function f, and actually proves something even stronger:

$$\text{(8)} \qquad \lim_{r\to 0+} \frac{1}{2r} \int_{x-r}^{x+r} |f(t) - f(x)|\, dt = 0 \qquad (f \in l^1(\mathbb{R}).$$

for each $x \in \mathbb{R}$ and f continuous with compact support.

Lebesgue's theorem says that if, instead of being continuous with compact support, f is merely *integrable on the real line*, then (8) holds for *almost every* $x \in \mathbb{R}$, with the same being true in higher dimensions, once we interpret the average on the left-hand side as extending over the ball of radius r centered at the point x.

More precisely, suppose $d \in \mathbb{N}$ and denote by $B_r(x)$ the open ball of radius $r > 0$, centered at $x \in \mathbb{R}^d$. Let μ denote Lebesgue measure for $\mathbb{R}^d$, with $L^1(\mu)$ (a.k.a. $L^1(\mathbb{R}^d)$) abbreviated to L^1. For later convenience we normalize μ to give measure 1 to balls of unit radius.

3.1 Theorem (Lebegue's Differentiation Theorem)**.** *For each $f \in L^1$:*

$$\text{(9)} \qquad \lim_{r\to 0+} \frac{1}{\mu\{B_r(x)\}} \int_{B_r(x)} |f - f(x)|\, d\mu = 0$$

for μ-almost-every $x \in \mathbb{R}^d$.

The *Lebesgue set* of $f \in L^1$ is the set of points $x \in \mathbb{R}^d$ for which (9) holds. Its complement has measure zero, and for each of its points x:

$$\lim_{r\to 0+} \frac{1}{\mu\{B_r(x)\}} \int_{B_r(x)} f\, d\mu = f(x)\,.$$

In fact, much more is true; here's an easy exercise that illustrates the power of the strong form (9) of Lebesgue's theorem. For simplicity, we state this only for dimension $d = 2$.

For the proof in all finite dimensions see, e.g. Rudin [23] §7.9, pp. 140-141.

3.2 Exercise. Suppose $\mathscr{R}$ is a family of closed rectangles in (for simplicity) $\mathbb{R}^2$ such that: (a) Each $x \in \mathbb{R}^2$ lies in a sequence of rectangles from $\mathscr{R}$ whose diameters converge to zero, and (b) there exist numbers $0 < c < C$ such that for each $R \in \mathscr{R}$, the ratio of the length of the larger side to the length of the smaller side lies between c and C. Let $d(R)$ denote the diameter of the rectangle R, and write $\mathscr{R}_x$ for those rectangles in $\mathscr{R}$ that contain the point x.

Show that for each $f \in L^1$ and each x in the Lebesgue set of f:

$$f(x) = \lim_{d(R)\to 0} \left\{ \frac{1}{\mu(R)} \int_R f \, d\mu : R \in \mathscr{R}_x \right\}.$$

3.3 Exercise. Show that the conclusion of Theorem 3.1 (hence also that of Exercise 3.2) continues to hold if we require only that f be *locally integrable* on $\mathbb{R}^d$, i.e., integrable over each ball.

Thus, for example, the Lebesgue Differentiation Theorem remains true for each space $L^p(\mathbb{R}^d)$, $(1 < p < \infty)$, none of which is contained within $L^1(\mathbb{R}^d)$.

To place Lebesgue's Theorem in the setting of Banach's program, we take the measure space to be euclidean d-space $\mathbb{R}^d$, with Lebesgue measure μ, the Banach space B to be $L^1 = L^1(\mathbb{R}^d)$, and the sequence[7] of maps (T_n) to be those given for $f \in L^1$ and $x \in \mathbb{R}^d$ by

$$T_n f(x) = \sup_{0<r<\frac{1}{n}} \frac{1}{\mu(B_r(x))} \int_{B_r(x)} |f - f(x)| \, d\mu \,, \tag{10}$$

noting that each of these maps is now *sublinear* (in the sense of sidenote 5 on page 5), and *Lebesgue measurable*.[8]

With this setup, the Lebesgue Differentiation Theorem, can be rephrased:

$$\lim_{n\to\infty} T_n f = 0 \;\; \textit{a.e. for each } f \in L^1.$$

For the dense subset of $B = L^1$ on which the desired a.e. convergence (9) holds, we take $D = C_c(\mathbb{R}^d)$, the space of continuous functions on $\mathbb{R}^d$ having compact support. That (9) holds for each $f \in C_c(\mathbb{R}^d)$ follows easily from the fact that each such function is uniformly continuous on $\mathbb{R}$. Thus, to complete Banach's program we need "only" an appropriate inequality for the maximal operator $T^\#$ defined by

$$T^\# f = \sup_n T_n f = \sup_{0<r<1} \frac{1}{\mu\{B_r(x)\}} \int_{B_r(x)} |f - f(x)| \, d\mu \,. \tag{11}$$

The argument we used in sidenote 8 to prove measurability for $T_n f$ works as well for both $T^\# f$, and the Hardy-Littlewood maximal function $\mathscr{M} f$ defined below.

The desired inequality will follow from:

3.4 Theorem (The Hardy-Littlewood Maximal Theorem)**.** *For $f \in L^1$ and $x \in \mathbb{R}^d$, define*

$$\mathscr{M} f(x) := \sup_{r>0} \frac{1}{\mu\{B_r(x)\}} \int_{B_r(x)} |f| \, d\mu \qquad (x \in \mathbb{R}^d). \tag{12}$$

Then for each $\lambda > 0$:

$$\mu\{\mathscr{M} f > \lambda\} \leq \frac{3^d}{\lambda} \|f\|_1 \,. \tag{13}$$

[7] We could as well work with the *one-parameter family* $\{T_r : r > 0\}$ of maps defined by the average on the right-hand side of (10). The proof of Theorem 2.1 goes through unchanged for this situation, and the same maximal theorem will apply.

[8] For $x \in \mathbb{R}^d$ we have $T_n(x) = \sup\{h_r(x) : 0 < r < 1/n\}$, where h_r is continous on $\mathbb{R}^d$. Suppose $T_n f(x) > \lambda$ for some $x \in \mathbb{R}^d$ and $\lambda > 0$. Then there is an index r such that $h_r(x) > \lambda$, so by continuity there is a neighborhood U of x on which $h_r > \lambda$, hence the same is true of $T_n f$. In other words, for every $\lambda > 0$ the set $\{T_n > \lambda\}$ is open, which establishes the measurability (in fact, the lower semi-continuity) of $T_n f$ on $\mathbb{R}^d$.

3.5 Corollary. *$\mathscr{M}f < \infty$ a.e. for each $f \in L^1$.*

Proof of Corollary. Fix $f \in L^1$. From inequality (13) we know that for each $n \in \mathbb{N}$ the set $\{\mathscr{M}f > n\}$ has measure $\leq 3^d\|f\|_1 \cdot \frac{1}{n}$. The set $\{\mathscr{M}f = \infty\}$ is the decreasing intersection of the sets $\{\mathscr{M}f > n\}$, and each of these sets lies in $\{\mathscr{M}f > 1\}$ which, by inequality (13), has finite measure, it follows from the "continuity of measure" that

$$\mu\{\mathscr{M}f = \infty\} = \lim_n \mu\{\mathscr{M}f > n\} = 0. \qquad \square$$

Inequality (13) calls to mind:

Chebyshev's Inequality. *For each $f \in L^1$ and $\lambda > 0$:*

$$\mu\{|f| > \lambda\} \leq \frac{1}{\lambda}\|f\|_1 .$$

Also known as "Markov's Inequality," or "The Chebyshev-Markov Inequality."
Proof. For $f \in L^1$:

$$\lambda\mu\{|f| > \lambda\} \leq \int_{\{|f|>\lambda\}} |f|\, d\mu \leq \|f\|_1.$$

This inequality suggests that a strategy for proving the Hardy-Littlewood Maximal Theorem might be to show that $f \in L^1$ implies $\mathscr{M}f \in L^1$, hopefully with a constant C (independent of f) such that $\|\mathscr{M}f\|_1 \leq C\|f\|_1$. Then Chebyshev's inequality would yield the Hardy-Littlewood Maximal Theorem (with C replacing 3^d on the right-hand side). Sadly:

3.6 Proposition. *If $f \in L^1$ is not a.e. equal to 0, then $\mathscr{M}f \notin L^1$.*

Proof. It's enough to prove this for $f = \chi_K$, the characteristic function of a compact subset K of $\mathbb{R}^d$ that has positive measure. Let δ be the radius of the smallest closed origin-centered ball that contains K. Fix $x \in \mathbb{R}^d \backslash K$. Then the closed ball of radius $r = |x| + \delta$, centered at x, contains K, so

$$(\mathscr{M}\chi_K)(x) \geq \frac{1}{\mu\{B_r(x)\}} \int_{B_r(x)} \chi_K\, d\mu = \frac{\mu(K)}{r^d} = \frac{\mu(K)}{(|x|+\delta)^d}.$$

Here we use the fact that an open ball and its closure have the same Lebesgue measure, as well as the scaling property of Lebesgue measure, and our normalization that balls of radius 1 have measure 1.

Since the right-hand side of this inequality is not integrable over $\mathbb{R}^d \backslash K$, neither is $\mathscr{M}\chi_K$. *Conclusion*: $\mathscr{M}\chi_K \notin L^1$. $\square$

Proof of the Hardy-Littlewood Maximal Theorem

Fix $\lambda > 0$ and $f \in L^1$ with $f \geq 0$ a.e. By the regularity of Lebesgue measure, it's enough to prove (13) for any compact subset K of $\{\mathscr{M}f > \lambda\}$. Fix such a set K.

For each $x \in K$ the definition of $\mathscr{M}f(x)$ promises that there is an open ball B_x centered at x such that

$$(14) \qquad \frac{1}{\mu\{B_x\}} \int_{B_x} f\, d\mu > \lambda\,, \quad \text{i.e., that} \quad \mu\{B_x\} < \frac{1}{\lambda} \int_{B_x} f\, d\mu\,.$$

Since the collection of balls $\mathscr{B} := \{B_x \colon x \in K\}$ is an open cover of K, there exists a finite subcollection $\{B_{x_j}\}_1^n$ that still covers K.

Let's write B_j for B_{x_j}, and (temporarily) indulge in some wishful thinking. Suppose our subcollection of B_j's were *pairwise disjoint*! Then we'd have:

$$\begin{aligned}
\mu\{K\} \le \mu\Big\{\bigcup_{j=1}^{n} B_j\Big\} &\le \sum_{j=1}^{n} \mu\{B_j\} \\
&\le \sum_{j=1}^{n} \frac{1}{\lambda}\int_{B_j} |f|\, d\mu \quad \text{[by (14)]} \\
&= \frac{1}{\lambda}\int_{\cup_j B_j} |f|\, d\mu \qquad \text{[by disjointness!]} \\
&\le \frac{1}{\lambda}\int_{\mathbb{R}} |f|\, d\mu = \frac{1}{\lambda}\|f\|_1
\end{aligned}$$

which would prove the Maximal Theorem (with a better constant).

Needless to say, the finite subcollection $\{B_j\}_1^n$ of K need *not* be pairwise disjoint. However all is not lost; a beautiful covering argument due to Norbert Wiener [27] saves the day! Choose a ball $B_j \in \mathscr{B}$ of largest radius (if there are several, make an arbitrary choice). Call this ball L_1, and remove from $\mathscr{B}$: not only L_1, but also all the "satellite" balls in $\mathscr{B}$ that intersect it. Note that $3L_1$, the open ball having the same center as L_1 and three times its radius, contains every one of these satellites!

If L_1 and its satellites exhaust all of $\mathscr{B}$, stop. Otherwise do the same with the remaining balls in $\mathscr{B}$, i.e., choose the largest one—call it L_2, remove it and all its satellites, and note that L_1 is disjoint from L_1, and that $3L_2$ swallows up all of L_2's satellites. Continue if necessary. Because the initial covering of K was finite, this process must eventually halt, resulting in a finite, pairwise-disjoint collection $\{L_k\}_1^\ell$ of open balls with centers in K, such that $\{3L_k\}_1^\ell$ covers K. Thus:

$$\mu(K) \;\le\; \mu\Big\{\bigcup_{k=1}^{\ell} 3L_k\Big\} \;\le\; \underbrace{\sum_{k=1}^{\ell}\mu\{3L_k\}}_{\text{since } K\subset\bigcup_k 3L_k} \;=\; \underbrace{3^d\sum_{k=1}^{\ell}\mu\{L_k\}}_{\text{since } \mu\{3L_k\}=3^d mu\{L_k\}} \quad .$$

The estimate now proceeds as in the "wishful thinking" argument, with L_k's replacing the B_j's. □

4 *Birkhoff's Ergodic Theorem.*

Our setting shifts now to $L^1 = L^1(\mu)$, where $(X, \mathscr{F}, \mu)$ is a *probability space*, i.e., a measure space for which $\mu(X) = 1$. We'll deal with mappings $\varphi\colon X \to X$ that are not just *measurable* (in that $\varphi^{-1}(F) \in \mathscr{F}$ for each $F \in \mathscr{F}$), but *measure-preserving* (in that $\mu(\varphi^{-1}F) = \mu(F)$

Whereupon we call μ a *probability measure*.

for every $F \in \mathscr{F}$). In this case the 4-tuple $(X, \mathscr{F}, \mu, \varphi)$ is called a *measure-preserving system*.

4.1 Example. Let $X = \{z \in \mathbb{C} \colon |z| = 1\}$, the unit circle of the complex plane, and $\mu =$ Lebesgue arc-length measure on X, normalized to have total mass 1. Let φ denote the "angle-doubling" map on X, defined by $Tz = z^2$. Being continuous, φ is a measurable mapping. φ is measure-preserving because $T^{-1}F$, for each not-too-large arc F of X, is the disjoint union of two arcs each having half the length of the original.

Suppose $(X, \mathscr{F}, \mu, \varphi)$ is a measure-preserving system. We study the orbits $(\varphi^n x)_0^\infty$ for each $x \in X$, where φ^n is the composition of φ with itself n times.[9] In particular, we study, for each $f \in L^1$, the "time averages"

[9] φ^0 is the identity map.

$$T_n f(x) = \frac{1}{n} \sum_{k=0}^{n-1} f(\varphi^n x) \qquad (x \in X). \tag{15}$$

Birkhoff [4], 1931. The term "ergodic", is said to have been invented by Boltzmann in the mid-1880's as a portmanteau of the Greek *ergon* (energy) and *hodos* (path).

The mathematical study of "Ergodic Theory" arises from "statistical mechanics," which arose from attempts by Boltzman and others in the late 1880's to explain thermodynamic phenomena via newtonian mechanics. For concise and informative discussions of this rich history, see Eisner, et al [7], especially Chapter 1, pp. 2–4, and Chapter 5, pp. 67-69.

4.2 Theorem (The Birkhoff Ergodic Theorem)**.** *For $(X, \mathscr{F}, \mu, \varphi)$ a measure-preserving system, and $f \in L^1$, there exists $\tilde{f} \in L^1$ such that:*

(a) $\lim_{n\to\infty} T_n f(x) = \tilde{f}(x)$ *for a.e.* $x \in X$.

(b) $\tilde{f} \circ \varphi = \tilde{f}$ *a.e.* .

(c) $\int \tilde{f}\, d\mu = \int f\, d\mu$

To set Birkhoff's Theorem within Banach's program, we already have: the measure space, the Banach space $B = L^1$, and the sequence (T_n) of linear maps. The "good" dense subset of B on which the desired a.e. convergence is known to hold will turn out to be $D = L^2 = L^2(\mu)$. The argument here, which is not entirely trivial, will make essential use of the Hilbert space structure of L^2; it will yield, as an additional benefit, the "Mean Ergodic Theorem" of John von Neumann.

The maximal inequality that will finish the proof of Birkhoff's result will arise from something that looks quite different:

4.3 Lemma (The "Maximal Ergodic Lemma")**.** *Suppose $(X, \mathscr{F}, \mu, \varphi)$ is a measure-preserving system, $f \in L^1$, and $T_n f$ is defined by* (15)*. Let*

$$T^{\#} f(x) = \sup_n T_n f(x) = \sup_n \frac{1}{n} \sum_{k=0}^{n-1} f(\varphi^n x) \qquad (x \in X).$$

Note that here, unlike Banach's definition of $T^{\#}$, *we do not* surround $T_n f$ by absolute values.

Then for each $f \in L^1$:

$$\int_{\{T^{\#} f > 0\}} f\, d\mu \;\geq\; 0\,.$$

Before proving this result, let's see how it implies the sort of maximal inequality required by Banach's Sufficient Condition (Theorem 2.1). Continuing with the notation of Lemma 4.3:

4.4 Theorem (The "Ergodic Maximal Theorem"). *For each $f \in L^1$:*

Rather than Banach's definition of $T^\#$, which uses $|T_n f|$, we use here $T_n|f|$, which gives an even larger version of the maximal function.

(16) $$\sup_{f \in L^1} \mu\{T^\#|f| > \lambda\|f\|_1\} \leq \frac{1}{\lambda} \qquad (\lambda > 0).$$

Proof. Given $f \in L^1$ and $\lambda > 0$, apply Lemma 4.3 to $g := |f| - \lambda$. Let

$$E_\lambda = \{T^\# g > 0\} = \{T^\#|f| > \lambda\}$$

Then

$$\begin{aligned} 0 \leq \int_{E_\lambda} g\, d\mu &= \int_{E_\lambda} (|f| - \lambda)\, d\mu \\ &= \int_{E_\lambda} |f|\, d\mu - \lambda\mu(E_\lambda) \\ &\leq \|f\|_1 - \lambda\mu(E_\lambda), \end{aligned}$$

which establishes (16). □

Proof of Theorem 4.2, part (a). By Theorem 4.4 the maximal inequality (2) holds for $T^\#$, so by Theorem **??** the "good set":

$$G := \{f \in L^1 : (T_n f)_1^\infty \text{ converges a.e.}\}$$

is closed in $L^1(\mu)$. It remains to show that G is dense in $L^1(\mu)$. For this, we'll exploit the Hilbert-space structure of L^2 to show that it has a dense subspace that belongs to G. Since L^2-convergence implies L^1 convergence G contains $L^2 = L^2(\mu)$, which is itself a dense subspace of L^1.

Since $\mu(X) < \infty$ we know that $L^2(\mu)$ is contained in $L^1(\mu)$, and since $\mu(X) = 1$ we know that $\|\cdot\|_1$ is $\leq \|\cdot\|_2$ on L^2. Since the simple functions form a dense subspace of L^2, they must also be dense in L^1.

To THIS END, we study the "Koopman Operator" U the the mapping φ induces on L^2:

$$Uf = f \circ \varphi \qquad (f \in L^2).$$

This operator was introduced into the study of Hamiltonian dynamical systems by Bernard Koopman [15, 1931], whose work inspired von Neumann's proof of his "Mean Ergodic Theorem" [22, 1932]. The survey article [18, 2021] describes some of the Koopman operator's recent theoretical and practical applications.

Since φ is measure-preserving, the change-of-variable formula insures that U is *isometric* on every L^p space, i.e., $\|Uf\|_p = \|f\|_p$ for each $f \in L^p$. We'll be particularly concerned with the subspace $\mathscr{K}$ of L^2 consisting of φ-invariant functions:

$$\mathscr{K} := \{f \in L^2 : f = f \circ \varphi\} = \ker(I - U)$$

For each $f \in \mathscr{K}$ and $n \geq 0$ we have $f \circ \varphi^n = f$, hence $T_n f = f$. Thus $\mathscr{K} \subset G$, trivially, so we can complete the proof that $L^2 \subset G$, by proving that the orthogonal complement $\mathscr{K}^\perp$ of $\mathscr{K}$ in L^2, also lies in G, i.e., that $(T_n f)$ converges a.e. for each f in $\mathscr{K}^\perp$.

Proof that $\mathscr{K}^\perp \subset G$. Because U is isometric we know that

(17) $$\ker(I - U) = \ker(I - U^*),$$

Proof of (17). Suppose $f \in \ker(I - U)$, i.e., that $Uf = f$. Then use the inner product of $L^2(\mu)$ to compute $\|f - U^*f\|_2$ where U^*, the adjoint of U while not necessarily an isometry is still a *contraction*: $\|U^*\| \leq 1$. The resulting computation (real scalars) is:

$$\begin{aligned} 0 = \|f - U^*f\|_2^2 &= \langle f - U^*f, f - U^*f\rangle \\ &= \|f\|_2^2 + \|U^*f\|_2^2 - 2\langle f, U^*f\rangle \\ &\leq 2\|f\|_2^2 - 2\langle Uf, f\rangle \\ &= 2(\|f\|_2^2 - \langle f, f\rangle) = 0 \end{aligned}$$ □

i.e., *if U fixes f then so does U^*.*

Thus, to finish our proof we need only show that $(T_n f)$ converges a.e. for each $f \in \operatorname{ran}(I-U)$. Fix $f \in \operatorname{ran}(I-U)$. Then for some $g \in L^2$:

$$f = (I-U)g = g - Ug = g - g\circ\varphi,$$

so

$$T_n f := \frac{1}{n}\sum_{k=0}^{n-1} f\circ\varphi^k = \frac{1}{n}\sum_{k=0}^{n-1}(g\circ\varphi^k - g\circ\varphi^{k+1}) = \frac{1}{n}g \; - \; \frac{1}{n}g\circ\varphi^n$$

Since $g/n \to 0$ a.e., all depends on showing that the same is true for $g\circ\varphi^n/n$. This is easy: $\|g\circ\varphi^n\|_2 = \|g\|_2$ since φ is measure-preserving, hence

$$\int\sum_n \left|\frac{g\circ\varphi^n}{n}\right|^2 d\mu \;=\; \sum_n\int\left|\frac{g\circ\varphi^n}{n}\right|^2 d\mu \;=\; \sum_n\left\|\frac{g\circ\varphi^n}{n}\right\|_2^2$$

$$= \;\sum_n\left\|\frac{g}{n}\right\|_2^2 \;=\; \|g\|_2^2\sum_n\frac{1}{n^2} \;<\; \infty$$

Consequently $\sum_n\left|\frac{g\circ\varphi^n}{n}\right|^2 < \infty$ a.e., so $\frac{g\circ\varphi^n}{n} \to 0$ a.e. . □

Conclusions. $(T_n f)_1^\infty$ converges a.e. for each f in a subspace of $L^2(\mu)$ that is dense in $L^2(\mu)$, hence also dense in $L^1(\mu)$. Thus we've fulfilled the conditions of Banach's two-step program, so have established that $(T_n f)_1^\infty$ converges a.e. for each $f \in L^1(\mu)$. This completes the proof of part (a) of the Birkhoff Ergodic Theorem. □

Proof of Theorem 4.2, parts (b) and (c). To prove part (b) we need only observe that for $f \in L^1(\mu)$

$$(T_n f)\circ\varphi + \tfrac{f}{n} = \tfrac{n+1}{n}T_{n+1}f \;\text{ a.e.},$$

where the left-hand side converges a.e. to $\tilde{f}\circ\varphi$ and the right-hand side a.e. to $\tilde{f}$.

As for part (c): It follows from part (a) and Fatou's Lemma that $\tilde{f} \in L^1(\mu)$. Recall that our argument that $(T_n f)$ converges a.e. on a dense subset of $L^1(\mu)$ actually showed that it converges on a dense subset of $L^2(\mu)$ to Pf, where P is the orthogonal projection taking $L^2(\mu)$ onto the kernel of the operator $U\colon f \to f\circ\varphi$. Since φ is measure-preserving for μ, the operator T_n is a contraction[10] on each of the Banach spaces $L^p(\mu)$ for $1 \le p \le \infty$. In particular, an "$\varepsilon/3$ argument" shows that:

$$T_n f \to Pf \textit{ in the norm of } L^2(\mu),$$

[10] Meaning: the norm of $T_n f$ is $\le$ that of f.

This is von Neumann's "Mean Ergodic Theorem" [22, 1932].

hence also that $T_n f \to Pf$ in the (weaker) norm of $L^1(\mu)$. In particular, $\int Pf\, d\mu = \int f\, d\mu$ for every $f \in L^1(\mu)$, thus completing the proof of Birkhoff's Ergodic Theorem with $\tilde{f} = Pf$ in part (c). □

REMARK. The orthogonal projection $P\colon L^2(\mu) \to \ker(I - U)$ is the Radon-Nikodym derivative of $f\, d\mu$ with respect to the restriction of μ to the σ-algebra $\mathscr{F}_\varphi$ of φ-invariant subsets of $\mathscr{F}$, i.e, it's the "conditional expectation" of f with respect to $\mathscr{F}_\varphi$, as is its extension to $L^1(\mu)$.

PROOF OF LEMMA 4.3 [11] For $f \in L^1(\mu)$ and $n \in \mathbb{N}$ let

$$S_n f \;:=\; nT_n = f \;+\; f \circ \varphi \;+\; f \circ \varphi^2 \;+\; \ldots \;+\; f \circ \varphi^{n-1}$$

[11] See, e.g., Taylor [26], Lemma 14.4, p. 196.

and write

$$S_n^\# f \;:=\; \max_{1 \le k \le n} S_k f\,.$$

It's enough to prove that for each $n \in \mathbb{N}$:

$$\int_{\{S_n^\# f > 0\}} f\, d\mu \ge 0\,. \tag{18}$$

To this end, note that for $1 \le k \le n$ and we have $(S_n^\# f)^+ \ge S_k f$, so

Notation: $g^+ := \max\{g, 0\}$

$$f + (S_n^\# f)^+ \circ \varphi \ge f + (S_k f) \circ \varphi = S_{k+1} f\,,$$

In particular, for $2 \le k \le n$

$$f \ge S_k f - (S_n^\# f)^+ \circ T\,,$$

and the same, trivially, for $n = 1$. Thus upon max-ing the right-hand side over $1 \le k \le n$ we obtain

$$f \ge S_n^\# f - (S_n^\# f)^+ \circ \varphi\,. \tag{19}$$

Finally, we integrate both sides of (19) over $E_n := \{S_n^\# f > 0\}$:

$$\begin{aligned}
\int_{E_n} f\, d\mu &\ge \int_{E_n} (S_n^\# f - (S_n^\# f)^+ \circ \varphi)\, d\mu \\
&= \int_{E_n} \big((S_n^\# f)^+ - (S_n^\# f)^+ \circ \varphi\big)\, d\mu && \text{By definition of } E_n. \\
&= \int_X (S_n^\# f)^+\, d\mu - \int_{E_n} (S_n^\# f)^+ \circ \varphi)\, d\mu && \text{b/c } (S_n^\# f)^+ = 0 \text{ on } X \backslash E_n. \\
&\ge \int_X (S_n^\# f)^+\, d\mu - \int_X (S_n^\# f)^+ \circ \varphi)\, d\mu && \text{Composition preserves positivity.} \\
&= \int_X (S_n^\# f)^+\, d\mu - \int_X (S_n^\# f)^+\, d\mu && T \text{ is } (\mu\text{-})\text{measure-preserving.} \\
&= 0
\end{aligned}$$

□

5 *Banach's Necessary Condition*

We now turn to Banach's First Principle, which asserts *necessity* of the uniform estimate (2) on page 2 for a.e. convergence. For this we return to the general setting:

- $(X, \mathscr{F}, \mu)$ is a measure space,
- M is the vector space of (μ-equivalence classes of) a.e. finite, $\mathscr{F}$-measurable, real-valued functions on X,
- B is a Banach space with norm $\|\cdot\|$, and
- T_n (for $n \in \mathbb{N}$) is a linear transformation $B \to M$ that is "continuous in measure."
- $T^\#$ is the maximal operator for the sequence (T_n):

$$T^\# f = \sup_{n \in \mathbb{N}} |T_n f| \qquad (f \in B).$$

- $\beta(\lambda) := \sup_{f \in B} \mu\{|T^\# f| > \lambda \|f\|\} = \sup_{\|f\|=1} \mu\{|T^\# f| > \lambda\} \quad (\lambda > 0).$

5.1 Theorem. *(Banach's necessary condition for a.e. convergence)*[12] *Suppose $\mu(X) < \infty$. If $T^\# f < \infty$ a.e. for each $f \in B$, then*

$$\lim_{\lambda \to \infty} \beta(\lambda) = 0\,. \tag{20}$$

[12] Banach [3], Theoréme I, page 356.

Convergence in measure is metrizable

To assist our proof, we note that the space M supports a complete metric d whose convergent sequences are precisely those that converge in measure.

5.2 Proposition. *For $f, g \in M$ let*

$$d(f,g) := \int \frac{|f-g|}{1+|f-g|}\, d\mu\,. \tag{21}$$

Then:

(a) *d is a translation-invariant metric on M.*

(b) *For a sequence (f_n) in M, and $f \in M$:*

$$f_n \to f \text{ in measure} \iff d(f_n, f) \to 0 \qquad (n \to \infty).$$

(c) *(M, d) is a complete metric space.*

Proof. (a) Let $\nu(f) := d(f,0)$ for $f \in M$. Then the triangle inequality for d is equivalent to

$$\nu(f+g) \le \nu(f) + \nu(g) \qquad (f,g \in M),$$

and this in turn follows from the fact that for each pair x, y of non-negative real numbers:

$$\frac{x+y}{1+x+y} \le \frac{x}{1+x} + \frac{y}{1+y}\,.$$

That the other metric axioms are satisfied by d is obvious.

(b) We may, without loss of generality, assume $f = 0$ (a.e.). Suppose $d(f_n, 0) \to 0$. Then for each $n \in \mathbb{N}$ and $\lambda > 0$:

$$d(f_n,0) \;=\; \int \frac{|f_n|}{1+|f_n|}\,d\mu \ge \int_{\{|f_n|>\lambda\}} \frac{|f_n|}{1+|f_n|}\,dm\,,$$

hence

$$d(f_n,0) \;\ge\; \frac{\lambda}{1+\lambda}\,\mu\{|f_n| > \lambda\}\,. \tag{22}$$

Here we use the monotone-increasing-ness of the function $t \to t/(1+t)$ on the positive real axis.

Thus $d(f_n, 0) \to 0 \implies \mu\{f_n > \lambda\} \to 0$ for each $\lambda > 0$, i.e., $f_n \to 0$ in measure.

Conversely, if $f_n \to 0$ in measure then some subsequence (f_{n_k}) converges to 0 a.e., so $d(f_{n_k}, 0) \to 0$ by the Dominated Convergence Theorem.

(c) We use the fact that a sequence in M that is Cauchy in measure must converge in measure[13]. Suppose a sequence in M is d-Cauchy. Then by inequality (22) our sequence is Cauchy in measure, hence convergent in measure to some $f \in M$. Thus by the just-proved part (b), our sequence d-converges to f. □

[13] See, e.g., Halmos [13], Theorem E, page 93, which uses "fundamental in measure" for what we're calling "Cauchy in measure".

For λ and ε both > 0, define

$$V_{\lambda,\varepsilon} := \{f \in M\colon \mu\{|f| > \lambda\} \le \varepsilon\}. \tag{23}$$

With this notation: $f_n \to f$ in measure iff for every positive λ and ε: eventually $f_n - f \in V_{\lambda,\varepsilon}$.

5.3 Proposition. *For each positive λ and ε the set $V_{\lambda,\varepsilon}$ is closed in (M, d).*

Proof. Suppose $f_n \in V_{\lambda,\varepsilon}$ for each $n \in \mathbb{N}$, and that $f_n \to f$ in measure. We wish to show that $f \in V_{\lambda,\varepsilon}$. To this end, fix $\nu > \lambda$ and note that

$$\{|f| > \nu\} \subset \{|f - f_n| > \nu - \lambda\} \cup \{|f_n| > \lambda\}\,,$$

hence

$$\mu\{|f| > \nu\} \le \mu\{|f - f_n| > \nu - \lambda\} + \mu\{|f_n| > \lambda\}$$

The first summand on the right $\to 0$ as $n \to \infty$ (since $f_n \to f$ in measure), while the second one is $\leq \varepsilon$ (since each $f_n \in V_{\varepsilon,\lambda}$). Thus $\mu\{|f| > \nu\} \leq \varepsilon$ for each $\nu > \lambda$, consequently $\mu\{|f| > \lambda\} \leq \varepsilon$, as desired. □

The same argument shows that

$$M \backslash V_{\lambda,\varepsilon} = \{f \in M \colon \mu\{|f| > \lambda\} \geq \varepsilon\}$$

is closed in M. Thus its complement

$$V^{o}_{\varepsilon,\lambda} := \{f \in M \colon \mu\{|f| > \lambda\} < \varepsilon\}$$

is open in M.

Proof of Theorem 5.1

Fix $f \in B$ and $\varepsilon > 0$, and note that, because $T^{\#} f < \infty$ a.e. there exists a positive integer $n = n(\varepsilon, f)$ such that $\mu\{T^{\#} f > n\} \leq \varepsilon$. Thus

$$B = \bigcup_{n\in\mathbb{N}} \{f \in B \colon \mu\{T^{\#} f > n\} \leq \varepsilon\} = \bigcup_{n\in\mathbb{N}} (T^{\#})^{-1}(V_{n,\varepsilon}) \tag{24}$$

For $N \in \mathbb{N}$ define $T^{\#}_N \colon B \to M$ by:

$$T^{\#}_N f = \sup_{1\leq n\leq N} |T_n f| \qquad (f \in B).$$

Since each T_n is a continuous map $B \to M$, the same is true of $T^{\#}_N$, hence $T^{\#}_N f \nearrow T^{\#} f$ as $N \nearrow \infty$, for each $f \in B$. Consequently:

$$\begin{aligned}(T^{\#})^{-1}(V_{n,\varepsilon}) &= \{f \in B \colon \mu\{T^{\#} f > n\} \leq \varepsilon\} \\ &= \bigcap_{N\in\mathbb{N}} \{f \in B \colon \mu\{T^{\#}_N f > n\} \leq \varepsilon\} \\ &= \bigcap_{N\in\mathbb{N}} (T^{\#}_N)^{-1}(V_{n,\varepsilon}) \,.\end{aligned}$$

Since the mapping $T^{\#}_N$ is continuous, and $V_{n,\varepsilon}$ is closed in M (Proposition 5.3) we know that $(T^{\#}_N)^{-1}(V_{n,\varepsilon})$ is closed in B, hence the same is true of $(T^{\#})^{-1}(V_{n,\varepsilon})$. Thus (24) expresses B as a countable union of these closed sets, so the Baire's Theorem[14] guarantees that one of them must contain a d-ball. Put more plainly: there exists $n \in \mathbb{N}$, $f_0 \in B$, and $\delta > 0$ such that

$$\|f - f_0\| \leq \delta \implies \mu\{A^{\#} f > n\} \leq \varepsilon \,. \tag{25}$$

[14] See, e.g., Axler [2], §6E, page 185.

It only remains to translate this "f_0-centered " result to the origin. To this end, write $g = (f - f_0)/\delta$. Thus g is an arbitrary unit vector in B, and upon noting that the additivity of each map T_n guarantees sub-additivity for $T^{\#}$, we obtain :

$$\delta\, T^{\#} g = T^{\#}(f - f_0) \leq T^{\#} f + T^{\#} f_0$$

which, along with (25) yields

$$\mu\{\delta T^{\#} g > 2n\} \leq \underbrace{\mu\{T^{\#} f > n\}}_{\leq\varepsilon \text{ since } \|f-f_0\|\leq\delta} + \underbrace{\mu\{T^{\#} f_0 > n\}}_{\leq\varepsilon} \leq 2\varepsilon$$

In summary, given $\varepsilon > 0$ there exists $n \in \mathbb{N}$ such that:

$$g \in B \text{ with } \|g\| = 1 \implies \mu\{T^{\#}g > 2n/\delta\} \leq 2\varepsilon\,. \tag{26}$$

Since can rewrite the definition of $\beta(\lambda)$ from (20) as

$$\beta(\lambda) = \sup_{\|g\|=1} \mu\{T^{\#}g > \lambda\}\,,$$

estimate (26) implies

$$\beta(2n/\delta) = \sup_{\|g\|\leq 1} \mu\{T^{\#}g > 2n/\delta\} \leq 2\varepsilon.$$

THUS: given $\varepsilon > 0$ we've found $\lambda_\varepsilon := 2n/\delta > 0$ such that $\beta(\lambda_\varepsilon) \leq 2\varepsilon$. Since the $\varphi(\lambda)$ decreases as λ increases, we have the desired result: $\lim\limits_{\lambda\to\infty} \beta(\lambda) = 0$. □

Estimates for $\beta(\lambda)$.

In each of our examples of Banach's program, the Banach space B has been the L^1 space of a measure,[15] and the resulting maximal inequality has had the form

[15] More specifically, a probability measure, or Lebesgue measure on $\mathbb{R}^d$.

$$\beta(\lambda) = O\left(\frac{1}{\lambda}\right) \qquad (\lambda \to \infty). \tag{27}$$

Elias Stein[16] showed that estimate (27) holds in Banach's necessary condition whenever B is a "nicely embedded" translation-invariant subspace of $L^1(G)$, where G is a compact abelian group, and each T_n is a bounded linear operator on B that commutes each translation operator $\tau_h\colon f \to f_h$.

[16] [25], 1961

Here "nicely embedded" means that $B \subset L^1(G)$ with the identity map a bounded operator $B \to L^1$. To say B is "translation-invariant" means that for each $f \in B$ and $h \in G$, the "h-translate" f_h, defined by

$$f_h(g) := f(g+h) \quad (g, h \in G)$$

is in B with $\|f\| = \|f_h\|$.

Stein showed that for the special case $B = L^p(G)$ with $1 \leq p \leq 2$, there results an even better improvement of Banach's necessary condition:

$$\beta(\lambda) = O\left(\frac{1}{\lambda}\right)^p \qquad (\lambda \to \infty). \tag{28}$$

Special cases of Stein's L^p result had been proved much earlier by Kolmogorov [14, 1925] and Calderon. Kolmogorov used an existing a.e. convergence result for conjugate Fourier series in L^1 to derive estimate (28) with $p = 1$ for the resulting maximal function, and Calderon showed that if the Fourier series of each L^2 function were to converge pointwise a.e., then the associated maximal function would obey (28) with $p = 2$.

Stein [25] attributes this result to Calderon. It appears (unattributed) along with a sketch of proof in [28] as Theorem 1.22 on page 65.

Part II

Hardy-Littlewood Redux

6 A Dirichlet Problem

In this section we will consider real-valued functions u on the euclidean upper half-plane UHP $= \{(x,y) \in \mathbb{R}^2 \colon y > 0\}$ that are *harmonic*, i.e., that satisfy *Laplace's equation*

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0 \tag{29}$$

In the real world, one might imagine UHP to be a flat metal sheet, to which, at time $t = 0$, an initial temperature $f(x)$ is applied at each point $x \in \mathbb{R}$. There results, at each point $(x,y) \in$ UHP, and each time $t > 0$, a temperature $U(x,y,t)$. At each point of UHP the "steady-state" temperature $u(x,y) = \lim_{t\to\infty} U(x,y,t)$, exists, and satisfies Laplace's equation.

at each point. The *Dirichlet problem* asks for a function u harmonic on UHP linked, in some way, to "boundary data" prescribed by an integrable function $f \colon \mathbb{R} \to \mathbb{R}$.

The Poisson Integral

Everything in this section springs from the function $p \colon \mathbb{R} \to \mathbb{R}$ defined by:

$$p(x) := \frac{1}{\pi}\frac{1}{1+x^2} \qquad (x \in \mathbb{R}), \tag{30}$$

where the factor $\frac{1}{\pi}$ is inserted to make the integral of φ over the real line equal to 1. The *Poisson kernel for level* $y > 0$ is the function P_y defined on $\mathbb{R}$ by

$$P_y(x) := \frac{1}{y}p\left(\frac{x}{y}\right) = \frac{1}{\pi}\frac{y}{x^2+y^2} \qquad (x,y) \in \text{UHP}. \tag{31}$$

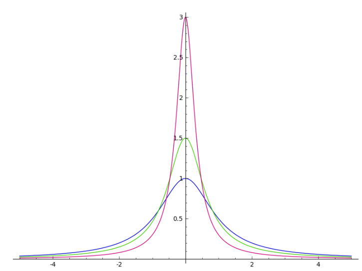

Figure 1: Graph of Poisson kernel $P_y(x) = \frac{1}{y}\varphi(x/y)$ for three different values of $y > 0$ (smaller values give "peakier" graphs).

Each Poisson kernel P_y is clearly non-negative, bounded, continuous on $\mathbb{R}$ and, for each $y > 0$, convergent to 0 as $x \to \pm\infty$. By a simple change of variable

$$\int_{\mathbb{R}} P_y(x)\,dx = \frac{1}{\pi}\int_{\mathbb{R}} \frac{dx}{1+x^2}\,dx = 1$$

for each $y > 0$.

6.1 Proposition. *For $f \in L^1(\mathbb{R})$, the "Poisson integral"*

$$P[f](x,y) := \int_{\mathbb{R}} P_y(x-t)f(t)\,dt \qquad (x \in \mathbb{R})$$

exists for every $x \in \mathbb{R}$, and defines a function $P[f]$ harmonic on UHP.

Proof. Upon identifying the point $(x,y) \in \mathbb{R}^2$ with the complex number $z = x + iy$, and observing that $P_y(x) = \text{Re}\left(i/(\pi z)\right)$ for each $z \in \mathbb{C}\backslash\{0\}$, we see that:

$$\pi P[f](z) := \int_{\mathbb{R}} \text{Re}\,\frac{i}{z-t}\,f(t)\,dt = \text{Re}\int_{\mathbb{R}} \frac{i}{z-t}\,f(t)\,dt\,,$$

The last equality uses from real-valuedness of f, and the definition of integral of a complex-valued function in terms of the integrals of its real and imaginary parts.

Thus to show that $P[f]$ is harmonic it will be enough to show that the function F defined on UHP by

$$F(z) := \int_{\mathbb{R}} \frac{f(t)}{z-t}\,dt \qquad (z \in \text{UHP})$$

The existence of the integral is not in doubt, since f is integrable and the integrand's denominator is bounded below by the distance from the (fixed) point $z \in$ UHP to the real line.

is analytic on UHP, i.e., that its complex derivative exists there.

To this end, fix points z and z_0 in UHP and compute:

$$\frac{F(z)-F(z_0)}{z-z_0} = -\int_{\mathbb{R}} \frac{f(t)}{(z-t)(z_0-t)}\,dt\,.$$

In the integral on the right, the integrand's denominator is bounded away from zero (with bound depending only on z and z_0), and as $z \to z_0$ the integrand converges pointwise (almost everywhere) to $f(t)/(z_0-t)^2$.

What we've really done here is justify an interchange of differentiation (with respect to z) with integration.

Thus by the Dominated Convergence Theorem, F is differentiable at z_0, with

$$F'(z_0) = -\int_{\mathbb{R}} \frac{f(t)}{(z-t)^2}\,dt\,.$$

Consequence: $P[f]$ is the real part of the iF, a function analytic on the upper half-plane, hence $P[f]$ is harmonic there. □

Thanks to the Cauchy-Riemann equations, the real part of any function analytic on an open subset of the plane is harmonic there (see, e.g. [24], Ch.2, §14-16, pp. 23–25).

Does $P[f]$ solve the Dirichlet problem for boundary data $f \in L^1$? The answer is "Yes" ... in the following sense:

6.2 Theorem. *If $f \in L^1$ and $u = P[f]$ then u is harmonic on* UHP, *and*

$$\lim_{y\to 0+} u(x,y) = f(x) \ \text{ for a.e. } x \in \mathbb{R}.$$

The harmonicity of u has already been established. To prove its a.e. convergence to f we'll follow Banach's program, taking

(a) $L^1 = L^1(\mathbb{R})$ for the Banach space B.

(b) The space of (a.e. equivalence classes of) Lebesgue-measurable real-valued functions on $\mathbb{R}$ for the space M.

(c) The one-parameter family sublinear maps T_y defined for $f \in L^1$ and $y > 0$ by

$$T_y f(x) = |P[f](x,y) - f(x)| \qquad (x \in \mathbb{R}).$$

In these terms the conclusion of Theorem 6.2 becomes: "$\lim_{y\to 0} T_y f(x) = 0$ for a.e. $x \in \mathbb{R}$."

(d) C_c = those $f\colon \mathbb{R} \to \mathbb{R}$ that are continuous and compactly supported, for the dense subset D of B on which Theorem 6.2 is known to hold. That C_c belongs to the "good set" for Theorem 6.2 is made clear by the Proposition below.

The denseness of D in L^1 follows easily from the fact that it contains each function continuous on $\mathbb{R}$ having compact support.

6.3 Proposition. *Suppose $f \in L^1(\mathbb{R})$ is bounded and uniformly continuous on $\mathbb{R}$. Then the function $u\colon \overline{\text{UHP}} \to \mathbb{R}$ defined by*

$\overline{\text{UHP}} := \{z = x + iy \in \mathbb{C} \colon y \geq 0\}$, the "closed upper half-plane."

$$u = \begin{cases} P[f] & \text{on} \quad \text{UHP} \\ f & \text{on} \quad \mathbb{R} \end{cases}$$

is harmonic on UHP *and continuous on* $\overline{\text{UHP}}$.

Proof. We'll need the following properties of the Poisson kernel:

(P1) $P_y \geq 0$ for each $y > 0$,

(P2) $\int_{\mathbb{R}} P_y(x)\,dx = 1$ for each $y > 0$, and

(P3) $\lim_{y\to 0+} \int_{|x|>\delta} P_y(x)\,dx = 0$ for each $\delta > 0$.

We noted the first two of these properties in the paragraph preceeding Proposition 6.1 above. For the third, fix positive numbers δ and y; then use the even-ness of P_y and the change-of-variable $x = ty$ to compute:

$$\int_{|x|>\delta} P_y(x)\,dx = 2\int_{x=\delta}^{\infty} \frac{1}{y}\varphi\Big(\frac{x}{y}\Big)\,dx = 2\int_{t=\delta/y}^{\infty} \varphi(t)\,dt\,.$$

Property (P3) follows from this and the integrability of φ over $\mathbb{R}$.

We prove the Proposition by showing that $\lim_{y\to 0+} P[f](x,y) = f(x)$, where the convergence is *uniform* for $x \in \mathbb{R}$. To this end, fix $\varepsilon > 0$ and use the uniform continuity of f to choose $\delta_0 > 0$ such that

$$|t| < \delta_0 \implies |f(x-t) - f(x)| < \varepsilon/2 \quad (\forall\, x \in \mathbb{R})\,.$$

By condition (P3) above, we can choose $\delta > 0$ so that

$$0 < y < \delta \implies \int_{|t|\geq\delta_0} P_y(t)\,dt < \frac{\varepsilon}{2\|f\|_\infty}\,.$$

$\|f\|_\infty := \sup_{t\in\mathbb{R}} |f(t)|\,.$

Fix $(x,y) \in$ UHP and note that:

$$\begin{aligned} P[f](x,y) - f(x) &= \int_{\mathbb{R}} P_y(x-t)f(t)\,dt - f(x) \\ &= \int_{\mathbb{R}} f(x-t)P_y(t)\,dt - f(x) && \text{[Change of variable]} \\ &= \int_{\mathbb{R}} [f(x-t) - f(x)]\,P_y(t)\,dt\,. && \text{[Property (P2) of Poisson kernel]} \end{aligned} \tag{32}$$

Now estimate (using in the first line the positivity of P_y):

$$\begin{aligned} |P[f](x,y) - f(x)| &\leq \int_{\mathbb{R}} |f(x-t) - f(x)|\,P_y(t)\,dt \\ &= \int_{|t|<\delta_0} + \int_{|t\geq\delta_0} |f(x-t) - f(x)|\,P_y(t)\,dt \\ &\equiv I_1 + I_2, \end{aligned}$$

where

$$I_1 = \int_{|t|<\delta_0} \underbrace{|f(x-t)-f(x)|}_{<\varepsilon \text{ by choice of } \delta_0} P_y(t)\,dt < \frac{\varepsilon}{2} \underbrace{\int_{|t|<\delta} P_y(t)\,dt}_{\leq 1 \text{ by } (P2)} \leq \frac{\varepsilon}{2}.$$

For the second integral we have this estimate for $0 < y < \delta$:

$$\begin{aligned} |I_2| &= \int_{|t|\geq\delta_0} \underbrace{|f(x-t)-f(x)|}_{\leq 2\|f\|_\infty} P_y(t)\,dt \\ &\leq 2\|f\|_\infty \underbrace{\int_{|t|\geq\delta_0} P_y(t)\,dt}_{<\frac{\varepsilon}{2\|f\|_\infty} \text{ by choice of } \delta} \\ &< \frac{\varepsilon}{2} \end{aligned}$$

Thus for each $x \in \mathbb{R}$:

$$0 < y < \delta \implies \big|P[f](x,y) - f(x)\big| \leq I_1 + I_2 < \varepsilon,$$

which establishes the uniform convergence of $P[f](x,y)$ to f. □

Thus the desired a.e. convergence will follow, via Banach's program (extended to one-parameter families of operators, as indicated in sidenote 7, page 8), from Theorem 2.1 and a favorable estimate on the maximal function $T^\# f = \sup_{y>0} T_y f$. For this we'll prove such an estimate for the *Poisson Maximal Function*

(33) $$P^\# f(x) := \sup_{y>0} \big(|P[f]|\big)(x,y) \qquad (x \in \mathbb{R}).$$

6.4 Theorem (The "Poisson Maximal Theorem"). *For each $f \in L^1(\mathbb{R})$ and $\lambda > 0$*

$$\mu\{P^\# f > \lambda\} \leq \frac{3}{\lambda}\|f\|_1,$$

where μ denotes Lebesgue measure on $\mathbb{R}$.

Proof that Theorem 6.4 implies Theorem 6.2. Since constant functions are fixed points for Poisson integrals, we have for each $f \in L^1$ and $x \in \mathbb{R}$:

$$\begin{aligned} (T^\# f)(x) &\leq \sup_{0<y<1} |P[f](x,y)| + |f(x)| \\ &\leq (P^\# f)(x) + |f(x)|. \end{aligned}$$

Thus, as in our proof of the Lebesgue Differentiation Theorem,

$$\begin{aligned} m\{T^\# f > \lambda\} &\leq m\{P^\# f > \lambda/2\} + m\{|f| > \lambda/2\} \\ &\leq \frac{6}{\lambda}\|f\|_1 + \frac{2}{\lambda}\|f\|_1 \\ &= \frac{8}{\lambda}\|f\|_1, \end{aligned}$$

which, thanks to Theorem 2.1, yields Theorem 6.2.

□

Proof of Theorem 6.4 (and therefore of Theorem 6.2). Fix $\gamma > 1$. We'll construct a "wedding-cake" $w\colon \mathbb{R} \to (0, \infty)$ whose graph lies between those of p and γp. To do this, draw the largest possible closed rectangle (sides parallel to the coordinate axes) that has top edge at height $p(0)$ and which fits between the graphs of p and γp. Now draw another such rectangle, the largest one whose top edge contains the bottom edge of the first rectangle, and which again fits between the two graphs. Continue the process (forever), obtaining a function $x \to w(x)$ whose graph looks like the side view of a wedding cake (see Figure 2).

Recall the Poisson kernel: $P_y(x) = y^{-1}p(x/y)$, where $p(x) = (\pi(1+x^2))^{-1}$.

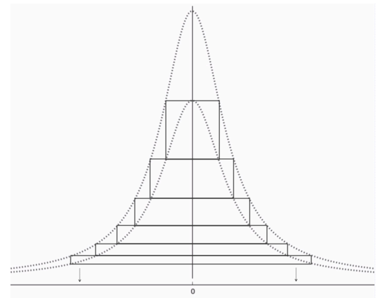

Figure 2: The "wedding-cake" construction

For $j \in \mathbb{N}$, denote by B_j the interval obtained by projecting the base of j-th rectangle formed above onto the horizontal axis, and let $H_j = \frac{1}{\mu(B_j)}\chi_{B_j}$. Thus $\int_{\mathbb{R}} H_j \, d\mu = 1$ for each $j \in \mathbb{N}$, and it's clear from its defining picture that there is a sequence (α_j) of positive constants such that $w = \sum_j \alpha_j H_j$. Moreover: $\sum_j \alpha_j \leq \gamma$.$^{(*)}$

χ_E denotes the characteristic function of the set $E \subset \mathbb{R}$ ($= 1$ on E, and 0 off E).

(*) *Proof*: $\gamma P_y \geq W_y$ and $\int_{\mathbb{R}} P_y \, dm = 1$, so

$$\gamma \geq \int_{\mathbb{R}} w \, dm = \sum_j \alpha_j \int_{\mathbb{R}} H_j \, dm = \sum_j \alpha_j.$$

For $(x, y) \in$ UHP let $W_y(x) = y^{-1}w(x/y)$, so that

$$P_y(x) \leq W_y(x) \leq \gamma P_y(x) \qquad (x \in \mathbb{R},\ y > 0).$$

Thus for $f \in L^1(\mathbb{R})$ and $x \in \mathbb{R}$:

$$\begin{aligned} P\big[|f|\big](x,y) &= \int_{\mathbb{R}} |f(x-t)| P_y(t) \, dt \\ &\leq \int_{\mathbb{R}} |f(x-t)| W_y(t) \, dt \\ &= \sum_j \alpha_j \int_{\mathbb{R}} f(x-t) H_j(t) \\ &= \sum_j \alpha_j \frac{1}{\mu(B_j)} \int_{B_j} |f(x-t)| \, dt \end{aligned}$$

Now make the substitution $s = x - t$ in the integrals in the last line above, and write $B_j(x)$ for $B_j + x$, the interval of the same size as B_j, but now with center x. The result is

$$P\big[|f|\big](x,y) \leq \sum_j \alpha_j \underbrace{\frac{1}{\mu(B_j(x))} \int_{B_j(x)} |f(s)| \, ds}_{\leq Mf(x)} .$$

where Mf is the Hardy-Littlewood Maximal Function (defined by (12) on page 8). *Conclusion*: For each $(x, y) \in$ UHP and each $f \in L^1(\mathbb{R})$:

$$P\big[|f|\big](x,y) \leq \gamma \, Mf(x) .$$

Upon taking the supremum of the left-hand side over $y > 0$ we obtain:

$$P^{\#}f(x) \leq \gamma Mf(x) \qquad \forall\ f \in L^1(\mathbb{R}) \text{ and } x \in \mathbb{R}.$$

Since this is true for each $\gamma > 1$ it's also true for $\gamma = 1$. Thus, for each $f \in L^1(\mathbb{R})$ and $\lambda > 0$:

$$\mu\{P^{\#}f > \lambda\} \;\leq\; \mu\{Mf > \lambda\} \;\leq\; \frac{3}{\lambda}\|f\|_1 .$$

with the final inequality provided by Theorem 3.4, the Hardy-Littlewood Maximal Theorem. This completes the proof of our Poisson Maximal Theorem, and with it, the proof of Theorem 6.2. □

L^p boundary data, $p > 1$.

We've just employed Banach's program to solve the Dirichlet problem of the upper half-plane (UHP) in the following sense (Theorem 6.2):

> *For each $f \in L^1(\mathbb{R})$ there exists a harmonic function u on* UHP *such that* $\lim_{y \to 0+} u(x,y) = f(x)$ *for a.e.* $x \in \mathbb{R}$.

The question arises:

> *If $p > 1$, does a similar result hold for $f \in L^p$?*

A straightforward application of Hölder's inequality shows that for each such f the Poisson integral exists and defines a function harmonic on UHP, so only the a.e. convergence is at issue.

A similar situation arose in our discussion of the Lebesgue differentiation theorem, proved initially for $f \in L^1$, then extended (via Exercise 3.3) to all f that are locally integrable, (and in particular, for $f \in L^p$ for each $p > 1$). The key to this generalization was the fact that the differentiation in question is *local*: the existence and value of

$$\lim_{r \to 0+} \frac{1}{\mu(B_r(x))} \int_{B_r(x)} f(t)\, dt$$

does not change no matter how the function f might be altered outside a neighborhood x.

The following lemma shows that, with a bit more work, we can establish the same kind of locality for the Poisson integral.

6.5 Lemma. *Suppose $1 < p \leq \infty$ and $f \in L^p$. Then for each $x \in \mathbb{R}$:*

$$\lim_{y \to 0+} \int_{|t-x|>1} f(t) P_y(x-t)\, dt = 0$$

Once this lemma has been established, we'll have:

6.6 Theorem. *If $f \in L^p(\mathbb{R})$ and $u = P[f]$ then:*

$$\lim_{y \to 0+} u(x,y) = f(x) \text{ for a.e. } x \in \mathbb{R}.$$

Proof that Lemma $\implies$ Theorem. Fix a unit-length interval I of the real line. For $f \in L^p$ let f_I coincide with f on I, and set it equal to 0 off I. Then f_I belongs to L^1 (Hölder's inequality), so by Theorem 6.2 we have, $\lim_{y\to 0+} P[f_I](x,y) = f_I(x)$ for a.e. $x \in \mathbb{R}$. In particular, this limit is $f(x)$ for a.e. $x \in I$. Now fix such an $x \in I$, and note that

$$P[f](x,y) = P[f_I](x,y) + \int_{|t-x|>1} f(t)P_y(x-t)\,dt$$

where (as $y \to 0+$) the first summand on the right converges to $f(x)$, and the second one to zero. Thus $\lim_{y\to 0+} P[f](x,y) = f(x)$ for a.e. $x \in I$.

The theorem follows upon decomposing the real line into countable a disjoint union of unit-length intervals, and applying the result of the above paragraph to each. □

Proof of Lemma. Fix $f \in L^p$. For $(x,y) \in$ UHP set

$$I_y(x) = \int_{|t-x|>1} f(t)P_y(x-t)\,dt$$

and let q denote the Hölder-conjugate index for p.

I.e., $q = 1$ if $p = \infty$, else $p^{-1} + q^{-1} = 1$.

$$\begin{aligned} |I_y(x)| &\le \int_{|t-x|>1} |f(t)|P_y(x-t)\,dt \\ &\le \|f\|_p \left(\int_{|x-t|>1} P_y(x-t)^q\,dt\right)^{1/q} \\ &= \|f\|_p \left(\int_{|t|>1} P_y(t)^q\,dt\right)^{1/q} \end{aligned}$$

where the second line follows from the first: obviously if $p = \infty$, otherwise by Hölder's inequality, and the third from the second by the obvious change-of variable (thanks to the even-ness of the Poisson kernel and the translation-invariance of Lebesgue measure). Now

$$\begin{aligned} 2\pi y^q \int_{t=1}^{\infty} P_y(t)^q\,dt &= \int_{t=1}^{\infty} \left(\frac{y^2}{y^2+t^2}\right)^q dt \\ &= y\int_{t=1/y}^{\infty} \left(\frac{1}{1+t^2}\right)^q dt \\ &\le\ y\int_{t=1/y}^{\infty} t^{-2q}\,dt \\ &= \tfrac{1}{2q-1}\,y^{2q}. \end{aligned}$$

Thus $|I_y(x)| = \mathrm{O}(y)$ as $y \to 0+$, which completes the proof of the Lemma. □

"Nontangential" convergence in the Dirichlet problem

We know from Theorem 6.2 that if $f \in L^1$ and $u = P[f]$, then for a.e. $x_0 \in \mathbb{R}$ the Poisson integral $u(x,y)$ of f converges to $f(x_0)$ as (x,y) converges to x_0 *along the vertical line* $x = x_0$. We'll see now that the arguments that produced this result can be easily modified to get a better "nontangential" result.

6.7 Definition. For $u\colon \text{UHP}] \to \mathbb{R}$ and points $x_0 \in \mathbb{R}$, to say $u(x,y)$ converges *nontangentially* to $w_0 \in \mathbb{R}$ means: For every $\alpha > 0$ and $\varepsilon > 0$ there exists $\delta = \delta(\alpha, \varepsilon) > 0$ such that $|u(x,y) - w_0| < \varepsilon$ whenever $|x - x_0| < \alpha y$ and $y < \delta$.

Geometrically, this means that for each given $\alpha > 0$ the value $u(x,y)$ approaches w_0 when (x,y) approaches $(x_0, 0)$ while staying within the "nontangential approach region"

$$\Omega_\alpha(x_0) = \{(x,y) \in \text{UHP}\colon |x - x_0| < \alpha y\}.$$

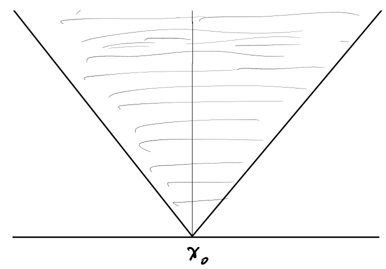

Figure 3: The approach region $\Omega_\alpha(x_0)$

6.8 Theorem. *If $f \in L^1$ then for a.e. $x_0 \in \mathbb{R}$, the Poisson integral $u = P[f]$ converges nontangentially to $f(x_0)$.*

Proof. Fix $\alpha > 0$, $x_0 \in \mathbb{R}$, and $f \in L^1$. For $y > 0$ define $T_y^\alpha\colon L^1 \to M$ by

$$T_y^\alpha f(x_0) = \sup_{|x - x_0| < \alpha y} |u(x,y) - f(x_0)|,$$

i.e, $T_y^\alpha f(x_0)$ is the supremum of $|P[f] - f(x_0)|$ over the intersection of $\Omega_\alpha(x_0)$ with the vertical line of height y. The conclusion of Theorem 6.8 can now be restated:

$$\lim_{y \to 0+} T_y^\alpha f(x_0) = 0 \quad \textit{for a.e. } x_0 \in \mathbb{R}. \tag{34}$$

Once again we have assembled the basic ingredients needed for Banach's program: The one-parameter family $\{T_y\colon y > 0\}$ of semilinear maps $L^1 \to M$ for which the desired convergence holds on a dense subset of L^1 (e.g., those integrable functions that are uniformly continuous on $\mathbb{R}$). Thus we can complete the proof of Theorem 6.8 by proving an appropriate inequality for the maximal function

$$T^\# f = \sup_{y > 0} T_y f \qquad (x_0 \in \mathbb{R},\ f \in L^1).$$

To this end, note that for $x_o \in \mathbb{R}$ and $y > 0$:

$$\begin{aligned} T_y^\alpha f(x_0) &= \sup_{|x - x_0| < \alpha y} \left| \int_{\mathbb{R}} P_y(x - t) f(t)\, dt - f(x_0) \right| \\ &\le \int_{\mathbb{R}} \sup_{|x - x_0| < \alpha y} |P_y(x - t) f(t)\, dt - f(x_0)|\, dt, \end{aligned}$$

so upon writing

$$P_y^{(\alpha)}(x_0) := \sup_{|x - x_0| < \alpha y} P_y(x) \tag{35}$$

we have

$$T_y^\alpha(x_0) \le \int_{\mathbb{R}} P_y^{(\alpha)}(x_0 - t)|f(t) - f(x_0)|\, dt \tag{36}$$

The rest of the proof will proceed exactly as for that of Theorem 6.2, once we establish

Lemma. *For each* $\alpha > 0$ *there exists a positive constant* C_α *such that*

$$P_y^{(\alpha)}(x_0) \le C_\alpha P_y(x_0) \qquad (x_0 \in \mathbb{R},\, y > 0). \tag{37}$$

For once this inequality is established, we'll know,

$$T_y^\alpha f(x_0) \le C_\alpha \int_{\mathbb{R}} P_y(x_0 - t)|f(t) - f(x_0)|\, dt$$

so that for each $x_0 \in \mathbb{R}$:

$$\begin{aligned} T^\# f(x_0) &\le C_\alpha P^\#(|f - f(x_0)|)(x_0) \\ &\le C_\alpha (M|f - f(x_0)|)(x_0) \end{aligned}$$

where $P^\#$ is the Poisson maximal function defined by (33) and M is the Hardy-Littlewood Maximal operator defined by (12) on page 8. Thus for each $f \in L^1$ and $\lambda > 0$

$$\begin{aligned} \mu\{T^\# f > \lambda\} &\le \mu\{M|f - f(x_0| > \lambda/C_\alpha\} \\ &\le \mu\{Mf > \lambda/(2C_\alpha)\} + \mu\{|f| > \lambda/(2C_\alpha)\} \\ &\le \frac{8C_\alpha}{\lambda}\|f\|_1 \end{aligned}$$

Thus the conditions of Banach's program are satisfied, so (modulo the proof of the Lemma) Theorem 6.8 is proven . □

Proof of the Lemma. We wish to show that

$$C_\alpha := \sup\left\{\frac{P_y^{(\alpha)}(x)}{P_y(x)} : x \in \mathbb{R},\, y > 0\right\} < \infty\,. \tag{38}$$

Recall that for each $y > 0$:

$$P_y(x) = \frac{1}{y}\, p\left(\frac{x}{y}\right) \qquad (x \in \mathbb{R})$$

where $p(x) = \dfrac{1}{\pi}\,\dfrac{1}{1+x^2}$.

In the same way:

$$P_y^{(\alpha)}(x) = \frac{1}{y}\, p^{(\alpha)}(\frac{x}{y}) \qquad (x \in \mathbb{R})$$

where

$$p^{(\alpha)}(x) = \sup\{(p(x-t) \colon |t| < \alpha\} \qquad (x \in \mathbb{R}).$$

Thus an easy calculation shows that our desired inequality (38) can be rewritten

(39) $$C_\alpha := \sup\left\{\frac{p^{(\alpha)}(x)}{p(x)} \colon x \in \mathbb{R},\, y > 0\right\} < \infty\,.$$

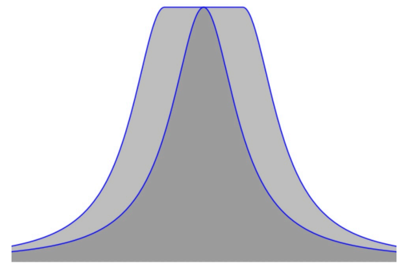
Figure 4: The graphs of p and $p^{(\alpha)}$

Figure 5 at the right shows that

$$p^{(\alpha)}(x) = \begin{cases} p(0) = \frac{1}{\pi}, & (|t| \le \alpha) \\ p(x-\alpha), & (0 \le t < \alpha) \\ p(x+\alpha), & (-\alpha < t \le 0) \end{cases}$$

Since $p^{(\alpha)}$ and p are both even functions, the supremum in (39) need only be computed for $x \ge 0$.

Now for $x > 0$:

(40) $$\frac{p^{(\alpha)}(x)}{p(x)} = \begin{cases} 1 + x^2 & (0 \le x \le \alpha) \\ \frac{x^2+1}{(x-\alpha)^2+1} & (x > \alpha) \end{cases}$$

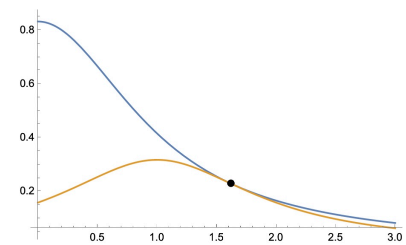

Figure 5: $C_\alpha p(x) \ge p(x-\alpha)$ for $\alpha = 1$ and $x \ge 0$. The dot is at the point of tangency.

and since $\frac{x^2+1}{(x-\alpha)^2+1} \to 1$ as $x \to \infty$, we see that the left-hand side of (39) is bounded on the non-negative real axis, so $C_\alpha < \infty$, which completes the proof of the Lemma (more precisely, a little Calculus shows that $C_\alpha = 1 + (\alpha/2)(\alpha + \sqrt{4+\alpha^2})$, see Figure 5). □

Comments and Complements

Higher dimensions. Let $\mathscr{U}$ denote the upper half-space of $\mathbb{R}^{d+1}$, i.e.,

$$\mathscr{U} := \{(x,y) \in \mathbb{R}^d \times \mathbb{R} \colon y > 0\}.$$

The analogue for $\mathscr{U}$ of the Poisson kernel for UHP is:

$$P_y(x) = \frac{1}{y^{n-1}}\varphi\left(\frac{x}{y}\right) \quad \text{where} \quad \varphi(x) = c_n \frac{1}{(1+|x|^2)^{n/2}}\,,$$

for $x \in \mathbb{R}^d$, $y > 0$ and c_n chosen to make $\int_{\mathbb{R}^d} \varphi\, dm = 1$. Then a change-of-variable in $\mathbb{R}^d$ shows—just as for the $n = 1$ case—that $\int_{\mathbb{R}^d} P_y\, dm = 1$ for each $y > 0$; in fact all the properties $P1 - P3$ for our original Poisson kernel [17] continue to hold for these higher dimensional ones.

Now m = Lebesgue measure on $\mathbb{R}^d$.

[17] See the proof of Theorem 6.3 for these.

We can then form the Poisson integral

$$P[f](x,y) := \int_{\mathbb{R}^d} P_y(x-t) f(t)\, dt \qquad ((x,y) \in \mathscr{U})$$

of any $f \in L^1(\mathbb{R}^d)$, verify that $P[f]$ is harmonic on $\mathscr{U}$ (now using an interchange of derivative and integral to verify that it satisfies Laplace's equation on $\mathscr{U}$), and proceed to prove by exactly the same steps as before that $P[f]$ solves the Dirichlet problem for $\mathscr{U}$, with a.e. convergence to boundary data $f \in L^1(\mathbb{R}^d)$.

See [2], Chapter 7, pp. 144-151 for the continuous and "mean" versions of the problem.

The higher dimensional analogue of the Poisson Maximal Theorem (Theorem 6.4) and therefore its resulting a.e. convergence theorems (Theorems 6.2 and 6.8) continue to hold, with essentially the same proofs—now based on the d-dimensional version of the Hardy-Littlewood Maximal Theorem.

7 *Fourier Series*

Our setting for this section shifts to the real interval $[-\pi, \pi]$, with μ denoting Lebesgue measure on that interval, normalized to have total mass 1, i.e., $d\mu(x) = \frac{dx}{2\pi}$.

Initially we'll work in (compex-valued) $L^2 = L^2(\mu)$, which is a Hilbert space with inner product

$$\langle f, g \rangle := \int f\,\overline{g}\,d\mu \qquad (f, g \in L^2),$$

and norm

$$\|f\|_2 := \sqrt{\langle f, f \rangle} \qquad (f \in L^2).$$

For $n \in \mathbb{Z}$ and $x \in \mathbb{R}$, let $e_n(x) = e^{inx}$. Then the bi-directional sequence $(e_n \colon n \in \mathbb{Z})$ is an orthonormal set in L^2 which, by the Weierstrass Approximation Theorem, is *complete* in that

$$\lim_{N\to\infty} \Big\| \sum_{|n|\le N} \langle f, e_n \rangle e_n - f \Big\|_2 = 0 \qquad (f \in L^2).$$

In other words, for each $f \in L^2$ the *Fourier series*

$$\sum_{n\in\mathbb{Z}} \langle f, e_n \rangle e_n \tag{41}$$

converges to f in the norm of L^2, in the sense that its sequence $(S_N)_0^\infty$ of *symmetric partial sums*

$$S_N f = \sum_{|n|\le N} \langle f, e_n \rangle e_n \tag{42}$$

In fact, the series (41)converges to f in L^2 *unconditionally*, i.e., no matter how its individual terms are rearranged.

converges in L^2 to f.

It's traditional write, $\hat{f}(n)$ for the n-th coefficient in the series (41):

$$\hat{f}(n) := \langle f, e_n \rangle = \int f(x)\,e^{-inx}\,d\mu(x) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\,e^{-inx}\,dx\,.$$

We call $\hat{f}(n)$ the *n-th Fourier coefficient* of f. The mapping $f \to \hat{f}$ is the *Fourier transform* which, by Parseval's theorem [REF] is a linear isometry taking L^2 *onto* the sequence space $\ell^2 = \ell^2(\mathbb{Z})$.

The notion of Fourier transform and Fourier series make sense, with the same definitions, for functions $f \in L^1 = L^1(\mu)$. Unfortunately the relationship between L^1 and its image under the Fourier transform is not as neatly described as in the L^2 case. Nevertheless, it's easily seen that the Fourier transform maps L^1 into the space $\ell^\infty = \ell^\infty(\mathbb{Z})$ of bounded (complex) sequences, and does so *contractively* in the sense that that $|\hat{f}(n)| \leq \|f\|_1$ for each $f \in L^1$. It's well known that:

(a) The Fourier transform maps L^1 one-to-one into (but *not* onto) the subspace c_0 of ℓ^∞ consisting of sequences that converge to 0 as $|n| \to \infty$, and

(b) There exist functions in L^1 whose Fourier series *do not* converge in L^1. [REF]

The problem of *pointwise convergence* of Fourier series dates back to the foundations of the subject. 19th century work of Dirichlet, Jordan, and du Bois Reymond showed that if f is continuous and 2π-periodic on the real line, *and of bounded variation*, then its Fourier series converges pointwise to f, but that the bounded-variation hypothesis could not be omitted.[18]

[18] See e.g., Rudin [23], §5.11–5.13, pp. 100-103 for the modern treatment of these matters.

Motivated by the Lebesgue theory of measure and integration, Nikolai Lusin asked in 1920 if, nevertheless, each such continuous function was the sum of its Fourier series at *almost every* point of the real line. Andrei Kolmogorov showed in 1923 (at age 20) that there exist functions $f \in L^1$ whose Fourier series *diverge* at almost every point of $\mathbb{R}$,[19] Nevertheless, Lusin's question remained unanswered until 1966, when the Swedish mathematician Lennart Carleson proved:

[19] Fundamenta Math. 4, pp. 324–328. A few years later Lusin produced an L^1-function whose Fourier series diverged at *every* point of $\mathbb{R}$.

Carleson's Theorem.[20] *If* $f \in L^2$, *then* $f(x) = \sum_{n\in\mathbb{Z}} \hat{f}(n)e^{inx}$ *for almost-every* $x \in [-\pi, \pi]$.

[20] Acta Math. 116 (1966) 135–157.

Carleson's proof proceeded (thanks to Theorem 1.3 and the density of the trigonometric polynomials) by establishing Banach's maximal condition (2) in the form

$$\sup_{f\in L^2} \mu\{T^\# f > \lambda\|f\|_2\} = \mathrm{O}\left(\frac{1}{\lambda^2}\right) \qquad (\lambda \to \infty),$$

where $T^\#$ is the "Fourier Maximal Operator"[21] defined for $f \in L^2$ and $|x| \leq \pi$ by:

[21] Also known as the "Carleson operator"

$$(T^\# f)(x) := \sup_N |S_N f(x)| = \sup_N \Big| \sum_{|n|\leq N} \hat{f}(n)e^{inx} \Big|.$$

Carleson's argument is challenging; subsequent authors have made the result more accessible (see, e.g., Lacey [19]), but it's still not for the faint-of-heart.

The fraught nature of the a.e. convergence problem for Fourier series lies in the integral representation of its symmetric partial sums:

$$S_N f(x) = \int f(x-t)\, D_N(t)\, d\mu(t) \qquad (f \in L^1)\,, \tag{43}$$

where D_N is the *Dirichlet kernel*

$$D_N(t) = \sum_{|n| \le N} e^{int} = \frac{\sin(N+\frac{1}{2})t}{\sin(\frac{t}{2})} \qquad (t \in \mathbb{R}).$$

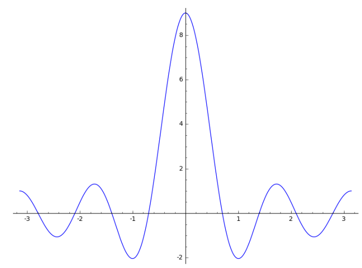

Figure 6: The Dirichlet Kernel D_4

In order for the representation (43) to make sense, we'll need to assume (now, and forever after) that each measurable function on $[-\pi, pi]$ has been 2π-periodically to the entire real line.
It turns out that $\lim_{N\to\infty} \|D_N\|_1 = \infty$, from which one can prove that both the operator S_N on L^1, and the linear functional $\varphi_{x,N} \colon f \to S_N f(x)$ on the continuous, 2π-periodic functions, have norms tending to ∞ as $N \to \infty$. In particular, the uniform boundedness principle insures that that there exists $f \in L^1$ whose Fourier series does not converge in L^1, and that there exist continuous periodic functions f whose Fourier series do not converge pointwise. [REFERENCES PLEASE!]

The situation for all modes of Fourier-series convergence improves considerably if, instead of focusing on the partial-sum sequence $(S_N)_0^\infty$, we instead focus on its sequence $(\sigma_N)_0^\infty$ of *arithmetic means*:

$$\sigma_N f = \frac{1}{N+1} \sum_{n=0}^{N} S_N f \qquad (f \in L^1)\,, \tag{44}$$

for which the integral representation is, for each $f \in L^1$:

$$\sigma_N f(x) = \int K_N(t) f(x-t)\, d\mu(t) \qquad (f \in L^1)\,. \tag{45}$$

Here K_N, called the *Fejér kernel*, is the N-th arithmetic mean of the Dirichlet-kernel sequence. The kindly nature of the sequence (σ_N) stems from its closed-form representation:

$$K_N(t) := \frac{1}{N+1} \left(\frac{\sin \frac{2N+1}{2} t}{\sin \frac{t}{2}} \right)^2 \tag{46}$$

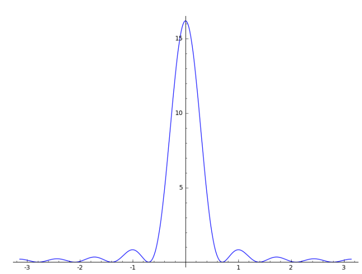

Figure 7: The Fejér kernel K_4

Thanks to this formula we see that the Fejer kernel has three crucial properties similar to those previously established here for the Poisson kernel (Properties P1–P3 on page 22) in the course of proving Proposition 6.3.

7.1 Lemma (Properties of the Fejer kernel)**.** *for* $N = 0, 1, 2, \ldots$ *and* $x \in \mathbb{R}$*:*

(a) $K_N(x) \geq 0$.

(b) $\int K_N(x)\, d\mu(x) = 1$ *for* $N = 0, 1, 2, \ldots$.

(c) $\lim_{N\to\infty} \max_{\delta \leq |x| \leq \pi} K_N(x) = 0$ *whenever* $0 < \delta < \pi$.

Proof. Property (a) is obvious from (46), while (b) follows immediately from the fact that $\int D_N(t)\, d\mu(t) = S_N 1(0) = 1$ for every N.

As for (c), we have from (46):

$$\max_{\delta \leq |x| < \pi} K_N(x) \leq \frac{1}{N+1} \cdot \frac{1}{\sin^2(\delta/2)} \to 0 \text{ as } N \to \infty. \qquad \square$$

Lemma 7.1 allows the argument we used in our initial solution of the Dirichlet Problem (Theorem 6.3, page 22) to work almost word-for-word in the present setting, resulting in:

7.2 Corollary. *If f is 2π-periodic and continuous on $\mathbb{R}$, then $\sigma_N f \to f$ uniformly on $\mathbb{R}$.*

i.e., $\lim_{N\to\infty} \max_{|x|\leq\pi} \left|(\sigma_N f)(x) - f(x)\right| = 0$.

By a *trigonometric polynomial* we mean a linear combination of exponentials e_n, i.e., a finite sum

$$p(x) = \sum_{n=1}^{N} a_n e^{inx} \qquad (N \in \mathbb{N},\ x \in \mathbb{R}).$$

7.3 Corollary (A Weierstrass Approximation Theorem)**.** *The trigonometric polynomials form a dense subspace of L^1.*

Proof. We know from Real Analysis that the continuous functions with compact support in the open interval $(-\pi, \pi)$ are dense in $L^1 = L^1([-\pi, \pi])$. Thus, if $f \in L^1$ and $n \in \mathbb{N}$, there exists a continuous function g with support in the open interval $(-\pi, \pi))$

$$\|f - g\|_1 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |f(x) - g(x)|\, dx \ < \ \frac{1}{2n}$$

Corollary 7.2 now provides a trigonometric polynomial p_n with $|g_n - p_n| < \frac{1}{2n}$ at each point of $[-\pi, \pi]$. Thus

$$\|f - p_n\|_1 \leq \|f - g_n\|_1 + \|g_n - p_n\|_1 < \frac{1}{2n} + \frac{1}{2n} = \frac{1}{n}.$$

Thus $\|f - p_n\| \to 0$ as $n \to \infty$. $\square$

7.4 Corollary. *For each $f \in L^1$:*

$$\lim_{N\to\infty} \|\sigma_N f - f\|_1 = 0.$$

Proof. GIVE THE THREE-EPSILON PROOF $\square$

The "Fejér Maximal Function

In the last section we used the "wedding cake construction," to showed how the Hardy-Littlewood maximal function dominates the Poisson maximal function. In fact, the argument works word-for-word if the Poisson kernel is replaced by for any "kernel" $J_y : \mathbb{R} \to (0, \infty)$ for $y > 0$, where J_y is non-negative, even, decreasing for $x > 0$, and integrable with $C := \sup_{y>0} \|J_y\|_1 < \infty$. If we define the operator T_y on L^1 by

$$T_y f(x) := \int_{\mathbb{R}} f(t) J_y(x-t)\, dt$$

then for each $f \in L^1$, the wedding-cake construction yields:

$$T^{\#} f(x) \le C\, \mathscr{M} f(x) \qquad (x \in \mathbb{R}),$$

where $\mathscr{M}$ is the Hardy-Littlewood Maximal Function.

In particular, (now returning to the setting of normalized Lebesgue measure on $[-\pi, \pi]$), note that the "central peak" of the graph of K_n sits on the interval $|t| \le \frac{2\pi}{2N+1}$, on which $K(t) \le K(0) = \frac{(2N+1)^2}{N+1}$. Define, for each $N \in \mathbb{N}$, the continuous function J_N to have value $K_N(0)$ on the support of that central peak, and to be an appropriate constant multiple $1/t^2$ on the rest of the interval $[-\pi, \pi]$. More precisely:

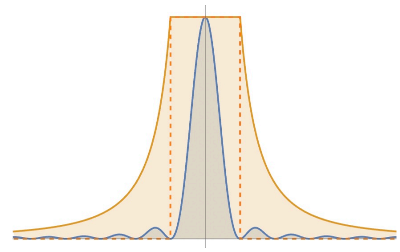

Figure 8: The Fejér kernel K_n and its majorant J_n

(47)
$$J_N(t) = \begin{cases} \frac{(2N+1)^2}{N+1} & \text{if} \quad |t| \le \frac{2\pi}{2N+1}, \\ \frac{4\pi^2}{N+1} \cdot \frac{1}{t^2} & \text{if} \quad \frac{2\pi}{2N+2} < |t| \le \pi\,. \end{cases}$$

7.5 Lemma. *For each $N \in \mathbb{N}$:*

(a) $J_N \ge K_N$ at each point of $[-\pi, \pi]$.

(b) $\|J_N\|_1 = \int_{-\pi}^{\pi} J_n(t)\, dt = \frac{8N}{N+1} < 8.$

Proof. Part (b) is a straightforward integration. For part (a), note that, by definition, the desired inequality is true for $|t| \le \frac{2\pi}{2N+1}$ (and is equality for $t = 0$). For the remaining values of t we have

$$\frac{K_N(t)}{J_N(t)} = \frac{t^2}{4\pi^2} \cdot \left(\frac{\sin((2N+1)\frac{t}{2})}{\sin(\frac{t}{2})} \right)^2 \le \frac{1}{4\pi^2} \cdot \frac{t^2}{\sin^2(\frac{t}{2})} \le \frac{1}{4\pi^2} \cdot \frac{t^2}{(t/\pi)^2}\,.$$

Since $\sin(t/2) \ge t/\pi$ on $[0, \pi]$.

In summary: K_N is $\le J_N$ on the support of the central peak of of K_N's graph, and is $\le J_N/4$ on the rest of the interval $[-\pi, \pi]$. □

Thanks to the “Wedding Cake construction” we know from part (b) of Lemma 7.5 that for each $f \in L^1$, $N \in \mathbb{N}$, and $|x| \leq \pi$:

$$\int J_N(t)|f(x-t)|\,d\mu(t) \leq 8\mathscr{M}f(x) \qquad (|x| \leq \pi),$$

so by part (a) of that Lemma:

$$\int K_N(t)|f(x-t)|\,d\mu(t) \leq 8\mathscr{M}f(x)\,. \tag{48}$$

7.6 Theorem (The “Fejer Maximal Theorem”). *For $f \in L^1$, define the “Fejer Maximal Function” $\sigma^\# f$ by:*

$$\sigma^\# f(x) = \Big|\sum_N \sigma_N f(x)\Big| \qquad (|x| \leq pi)$$

Then $\sigma^\# f \leq 8\mathscr{M}f$ on $[-\pi, \pi]$.

7.7 Corollary. $\lim_{N\to\infty} \sigma_N f = f$ *a.e. for each* $f \in L^1$.

8 References

1. Sheldon Axler, *Measure, Integration, and Real Analysis*,[22] Springer 2020. Open access: downloadable at `https://measure.axler.net`

2. Sheldon Axler, Paul Bourdon, and Wade Ramey, *Harmonic Function Theory*, 2nd ed., Springer 2001.

3. Stefan Banach, *Sur la convergence presque partout de fonctionelles linéares*, Bull. Sci. Math. 50 (1926) 27–32 & 36–43

4. George D. Birkhoff, *Proof of the ergodic theorem*, Proc. Nat. Acad. Sci. USA 17 (1931) 656–660.

5. Arlen Brown, P.R. Halmos, and A. L. Shields, *Cesàro operators*, Acta Sci. Math. (Szeged) 26 (1965) 125-137.

6. Lennart Carleson, On convergence and growth of partial sums of Fourier series, Acta Mathematica, 116 (1965) 135–157.

7. T. Eisner, B. Farkas, M. Haase, R. Nagel, *Operator Theoretic Aspects of Ergodic Theory*, Springer 2015. Draft copy freely available at: `https://www.math.uni-leipzig.de/~eisner/`

8. Adriano Garsia, *Topics in Almost Everywhere Convergence*, Markham 1970.

9. G. H. Hardy, *Note on a theorem of Hilbert*, Math. Z. (1920) 314–317. `http://gdz.sub.uni-goettingen.de/dms/load/img/?PID=GDZPPN002365391`

10. G. H. Hardy and J.E. Littlewood, *Some new properties of Fourier constants*, Math. Ann. 97 (1926) 159–209

11. G. H. Hardy and J. E. Littlewood, *A maximal theorem with function theoretic applications*, Acta Math. 54 (1930) 81–116.

12. G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities*, 2nd ed., Cambridge Univ. Press, 1952.

13. Paul R. Halmos *Measure Theory*, Springer 1974 (originally published by van Nostrand, 1950).

14. Andrei Kolmogorov *Sur les fonctions harmonique conjuguées et les séries de Fourier*, Fundamenta Math. 7 (1925) 23–28.

15. Bernard O. Koopman, *Hamiltonian systems and transformation in Hilbert space*, Proc. Natl. Acad. Sci. USA 17 (1931), 315–318.

16. Bernard Koopman and John von Neumann, *Dynamical systems of continuous spectra*, Proc. Nat. Acad. Sci. USA, 18 (1932), 255—266.

[22] See Chapter 4 for the maximal-function proof of the Lebesgue differentiation theorem.

17. Alois Kufner, Lech Maligranda, and Lars-Erik Persson, *The prehistory of the Hardy Inequality*, Amer. Math. Monthly 113 (2006) 715–732.

18. Igor Mezić, *Koopman operator, geometry, and learning of dynamical systems*, Notices Amer. Math. Soc. 68 (2021), 1087–1105.

19. Michael T. Lacey, *Carleson's Theorem, proof, complements, variations*, Publicacions Matemàtiques, 48 (2004) 251–307. See also arXiv:math/0307008 for an expanded version.

20. David Minda, *The Dirichlet problem for a disk*, Amer. Math. Monthly 97 (1990) 220–223. Free download at `http://www.jstor.org/stable/2324689`.

21. Calvin C. Moore, *Ergodic theorem, ergodic theory, and statistical mechanics*, Proc. Nat. Acad. Sci. USA, 112 (2015), 1907–1911

22. John von Neumann, *Proof of the quasi-ergodic hypothesis*, Proc. Nat. Acad. Sci. USA 18 (1932) 70—82.

23. Walter Rudin, *Real and Complex Analysis*, 3rd ed.,[23] McGraw-Hill 1987

[23] See Chapter 7 for the maximal-function proof of the $\mathbb{R}^n$ version of the Lebesgue Differentiation Theorem.

24. Donald Sarason, *Complex Function Theory*, 2nd ed. American Math. Society 2007.

25. Elias M. Stein, *On limits of sequences of operators*, Annals of Math. 74 (1961) 140–170.

26. Michael Taylor, *Chapter 14, Ergodic Theory*, University of North Carolina Lecture Notes, available at:
`http://www.unc.edu/math/Faculty/met/measch14.pdf`

27. Norbert Wiener, *The ergodic theorem*, Duke Math J. 5 (1939) 1–18.

28. Antoni Zygmund, Trigonometric Series, Vol. 2, Cambridge University Press 1959.